Ciencias Cognitivas

Daniel C. Dennett

LA *LIBERTAD* de *ACCION*

Un análisis de la exigencia de libre albedrío

gedisa editorial

Daniel C. Dennett

LA LIBERTAD DE ACCION

Serie CLA•DE•MA
Grupo: CIENCIA COGNITIVA

Editorial Gedisa ofrece
los siguientes títulos sobre

CIENCIAS COGNITIVAS

pertenecientes a sus diferentes
colecciones y series
(Grupo "Psicología")

F. VARELA, E. THOMPSON Y E. ROSCH	*El mundo en persona*
DANIEL C. DENNETT	*La libertad de acción*
HILARY PUTNAM	*Representación y realidad*
DANIEL C. DENNETT	*La actitud intencional*
FRANCISCO J. VARELA	*Conocer*
JEROME BRUNER	*Realidad mental y mundos posibles*

LA LIBERTAD DE ACCIÓN

Un análisis de la exigencia de libre albedrío

por

Daniel C. Dennett

gedisa
editorial

Título del original en inglés:
The Elbow Room
Originally published in English by Oxford University Press
© *by* Daniel C. Dennett, 1984

Traducción: Gabriela Ventureira
Diseño de la cubierta: Gustavo Macri

Asesores: Sergio Balari, Nuria Bel,
Elizabeth Gilboy, Nuria Sebastián, Olga Soler y José M.ª Sagrera
Coordinador: Carlos Clavería

Primera edición, Barcelona, España, 1992

Derechos para todas las ediciones en castellano

© *by* Editorial Gedisa, S. A.
Muntaner, 460, entlo., 1.ª
Tel. 201 60 00
08006 - Barcelona, España

ISBN: 84-7432-426-2
Depósito legal: B. 8.541-1992

Impreso en España
Printed in Spain

Impreso en Libergraf,
c/. Constitució, 19 - 08014 Barcelona

Queda prohibida la reproducción total o parcial por cualquier medio de impresión, en forma idéntica, extractada o modificada, en castellano o cualquier otro idioma.

ADVERTENCIA

ESTA ES UNA COPIA PRIVADA PARA FINES EXCLUSIVAMENTE EDUCACIONALES

QUEDA PROHIBIDA
LA VENTA, DISTRIBUCIÓN Y COMERCIALIZACIÓN

- El objeto de la biblioteca es facilitar y fomentar la educación otorgando préstamos gratuitos de libros a personas de los sectores más desposeídos de la sociedad que por motivos económicos, de situación geográfica o discapacidades físicas no tienen posibilidad para acceder a bibliotecas públicas, universitarias o gubernamentales. En consecuencia, una vez leído este libro se considera vencido el préstamo del mismo y deberá ser destruido. No hacerlo, usted, se hace responsable de los perjuicios que deriven de tal incumplimiento.
- Si usted puede financiar el libro, le recomendamos que lo compre en cualquier librería de su país.
- Este proyecto no obtiene ningún tipo de beneficio económico ni directa ni indirectamente.
- Si las leyes de su país no permiten este tipo de préstamo, absténgase de hacer uso de esta biblioteca virtual.

"Quién recibe una idea de mí, recibe instrucción sin disminuir la mía; igual que quién enciende su vela con la mía, recibe luz sin que yo quede a oscuras" ,

—Thomas Jefferson

COPIA PRIVADA

Para otras publicaciones visite
www.lecturasinegoismo.com
Facebook: Lectura sin Egoísmo
Twitter: @LectSinEgo
o en su defecto escríbanos a:
lecturasinegoismo@gmail.com
Referencia: 779

Dedicado
a la memoria
de Gilbert Ryle

Indice

1. Por favor, no alimente a los ogros 13

1. El eterno y cautivante problema 13
2. Los hombres malos 19
3. La avispa *sphex* y otras preocupaciones 23
4. Resumen 30

2. Hacer la razón práctica 34

1. ¿De dónde provienen las razones? 34
2. Motores semánticos, máquinas de movimiento perpetuo y un generador de intuición defectuoso 41
3. Reflexión, lenguaje y conciencia 49
4. Comunidad, comunicación y trascendencia 59

3. Control y autocontrol 66

1. "Debido a circunstancias que están fuera de nuestro control" 66
2. Control simple y autocontrol simple 68
3. Control sin agente y el concepto de causación 73
4. Agentes en competencia 77
5. Los usos del desorden 83
6. "¡Suéltese!" 86

4. El yo que se hace a sí mismo 91

1. El problema del yo que desaparece 91
2. El arte de la autodefinición 98
3. Probar suerte 110
4. Resumen 119

5. Actuar según la idea de libertad 120

1. Cómo seguir deliberando en los tiempos que corren 120
2. El diseño del deliberador perfecto 127
3. Oportunidades reales 134
4. Evitar, evitable, inevitable 142

6. "Podría haber obrado de otra manera" 151

1. La importancia de haber podido obrar de otra manera 151
2. Las cosas que nos interesan 160
3. El verbo "poder" 165

7. La necesidad del libre albedrío 175

1. La refutación del nihilismo 175
2. La responsabilidad disminuida y el espectro de la exculpación creciente 179
3. La negación del secreto terrible 188

BIBLIOGRAFÍA 197

INDICE TEMÁTICO 211

1

Por favor, no alimente a los ogros

El filósofo es aquel que,
en un simposio sobre la pena capital,
presentará un trabajo sobre la paradoja del ahorcado.
James D. McCawley

1. El eterno y cautivante problema

La idea de destino es más antigua que la filosofía. Desde sus comienzos, los filósofos trataron de mostrar por qué es errónea la creencia de que nuestra suerte está sellada desde antes del nacimiento. Se consideró de suma importancia demostrar que no representamos meramente nuestro destino sino que, en cierta medida, elegimos nuestros propios cursos de acción al tomar decisiones y al no permitir que las "decisiones" nos sucedan.

En la temprana filosofía griega, las ideas acerca de la causalidad constituían el foco de atención; a alguien se le ocurrió preguntarse si todos los acontecimientos físicos eran causados o estaban determinados por la suma total de los acontecimientos anteriores. En ese caso, es decir, si el *determinismo* es verdadero, nuestras acciones en tanto que acontecimientos físicos deben estar determinadas. Si el determinismo es verdadero, entonces, cada acto o decisión sería el resultado inexorable de la suma de fuerzas físicas que actúan en ese instante, que a su vez es el resultado de la suma de fuerzas físicas que actuaron en el instante anterior y así sucesivamente, hasta el principio del tiempo.

¿Cómo es posible, entonces, la libertad? Los epicúreos, que sostenían un *materialismo* asombrosamente moderno (creían que

la mente, al igual que todas las cosas, estaba compuesta por átomos materiales), trataron de librarse de la pesadilla de la elección predestinada rompiendo la trama de la causalidad universal. Postulaban que los átomos mostraban ocasionalmente "desvíos casuales".

> Además, si todos los movimientos están conectados entre sí, si lo nuevo surge de lo viejo en un orden determinado —si los átomos jamás se desvían para originar un nuevo movimiento que rompa las ataduras del destino, la perpetua secuencia de causa y efecto—, ¿cuál es la fuente del libre albedrío que poseen las cosas vivientes de la Tierra? (Lucrecio, *La naturaleza de las cosas*, II, líneas 250-255.)

La dificultad, tantas veces señalada, de esta solución y de sus variantes más modernas reside en que tales desvíos casuales, si acaso suceden, no parecen ser capaces de proporcionar la clase de libre albedrío que deseamos. Si un átomo de mi cerebro vira de pronto debido a un desvío casual, no lo hace "por alguna razón", y si la consecuencia de ello es que yo haga o decida algo importante, entonces estoy por completo a merced de los desvíos casuales. Una elección casual, tan ciega y arbitraria como una tirada de dados o un giro de la rueda de la fortuna, es tan indeseable como una elección determinada de antemano. En verdad, son muchos los que han pensado que es *más* indeseable y han propuesto otras formas de *conciliación* (o, como se dice con más frecuencia, de *compatibilización*) entre el libre albedrío y el determinismo. Algunos de estos intentos de conciliación son apenas menos deplorables que la concepción que, supuestamente, deben atacar.

Los estoicos, por ejemplo, aseguraban que es posible gozar de cierta libertad si, en lugar de luchar contra lo inevitable, adecuamos nuestros deseos a las circunstancias, e instaban a adoptar una actitud de sabia resignación que denominaron *apatheia*. Si bien es cierto que en su periplo etimológico el concepto de *apatheia* se ha simplificado y empobrecido hasta convertirse en el actual concepto de *apatía*, también es cierto que a los filósofos estoicos les encantaba explicar su doctrina valiéndose de metáforas muy deprimentes. Afirmaban que a cada uno de nosotros le es asignado un papel a desempeñar en la tragedia de la vida y que lo único que podemos hacer es decir la parte que nos corresponde de la mejor manera posible. No hay ningún espacio para el *ad libitum*. Imaginemos un perro atado con una soga a la parte trasera de un carro. El perro puede trotar pacíficamente detrás del carro o bien resistirse. En ambos casos llegará al mismo destino, pero si se resigna a su suerte y saca el mayor provecho del viaje, gozará de una cierta libertad (ser conducido por la vida con una soga al cuello, ¡vaya libertad!).

Por más de dos milenios, los filósofos han tratado de crear una doctrina del libre albedrío que sea a la vez más atractiva y más racional que las desalentadoras doctrinas de los comienzos. Se dice con frecuencia (y, aunque esto es plausible, me pregunto hasta qué punto es exacto) que se ha escrito mucho más sobre el libre albedrío que sobre cualquier otro tema filosófico. Los filósofos deberían sentirse por lo menos molestos, ya que a pesar de tan denodados esfuerzos, el progreso ha sido mínimo.

Se dice, también, que la filosofía no es una ciencia y que ahí reside el problema. Si fuera una ciencia, solucionaría los problemas que tienen solución y disolvería o descartaría el resto.

Otros opinan que el problema de la filosofía es que ha intentado ser "científica" en materias que sólo pueden ser tratadas por el arte. Si abandonara sus amoríos con el método científico, ya no tendría que plantear su temática en términos que garantizan el fracaso. Pienso que el problema de la filosofía es más complejo de lo que suponen el científico y el artista pues comparte —y debe compartir— las aspiraciones y los métodos de ambos.

Hay preguntas filosóficas ineludibles —la pregunta por el libre albedrío es una de ellas— que exigen respuestas claras, consistentes y bien fundadas. No debemos permitir que nos sobornen con respuestas alusivas, impresionistas, por muy interesantes y conmovedoras que sean. Sin embargo, la mayoría de los intentos de tratar las preguntas filosóficas con verdadero rigor —y las preguntas sobre el libre albedrío no son la excepción— se topan con los problemas de la formalización prematura. Hay gran abundancia de trabajos técnicos sobre el libre albedrío que poseen, irónicamente, un mero interés estético (para los *connaisseurs* de las fórmulas arquitectónicas o de la sofistería lógica), pues no logran ponerse en contacto con los verdaderos problemas. La primera cuestión a resolver en filosofía es, desde siempre, la de encontrar el método apropiado de investigación, y nunca hubo consenso en cuanto al método correcto. Todas las obras sobre el libre albedrío hacen una declaración, *ipso facto,* acerca del método y de cómo debe encararse el problema. En este libro, plantearemos la cuestión del método con menos solemnidad. Mi método toma a la ciencia con toda seriedad, pero sus tácticas se parecen mucho más a las del arte.

En mis días de estudiante, pensaba convertirme en escultor y me dedicaba con más fervor a los bloques de piedra o de madera que a la filosofía. Mientras trabajaba en este libro, me di cuenta de que nunca había abandonado los métodos aprendidos en el taller: simplemente, había cambiado el medio de comunicación. A diferencia del dibujante, que de un solo trazo debe lograr una línea perfecta, el escultor se da el lujo de picar y pulir hasta que líneas y superficies

parezcan perfectas. En primer término, se desbasta el bloque mientras se retrocede y se lo mira al sesgo para asegurarse de que uno se está acercando al producto final, oscuramente entrevisto. Cuando la pieza ha alcanzado las proporciones correctas, se vuelve a cada una de las superficies todavía sin pulir y sólo entonces se esfuerza uno en lograr los detalles más finos.

Algunos filósofos no sienten ninguna afinidad con este método cuando lo encuentran en filosofía. Se impacientan ante las soluciones meramente desbastadas pues sólo desean ver aristas netas y bien pulidas desde el comienzo. Al igual que ellos, aspiro a un producto bien terminado, pero cuestiono su estrategia. ¡Es tan difícil empezar con el pie derecho en filosofía y los riesgos de esa estrategia se hacen tan evidentes en la literatura sobre el libre albedrío, plagada de brillantes pero inútiles fragmentos! Se ha progresado muy poco en la cuestión del libre albedrío (y ése es uno de los temas del libro) pues los filósofos corren presurosos a ocuparse de lo que consideran las partes más importantes del problema concediéndoles una atención algo miope y no logran percibir la forma del aspecto principal del problema. Del opaco bloque de mármol emergen el rostro, las manos y los pies exquisitamente pulidos, pero no se ha dejado ningún espacio para los codos. Buena parte del trabajo de este libro consiste en desbastar las formas de las partes que los filósofos generalmente se apresuran a omitir, luego de palmaditas y promesas.

A menudo se señala que el problema del libre albedrío es el único problema filosófico interesante e incluso cautivante: personas que por lo general no se interesan en la filosofía se preocupan, sin embargo, por el problema y se sienten verdaderamente perturbadas ante la perspectiva de que las respuestas a sus preguntas puedan resultar equivocadas.

¿Por qué las personas encuentran cautivante el problema del libre albedrío? En parte, porque atañe profundamente a nuestra situación en el universo y a "nuestra condición humana", como se dice pomposamente. Pero también, agregaría, porque los filósofos han invocado una hueste de horribles ogros y luego han afirmado, de un modo subliminal y totalmente ílícito, que el problema del libre albedrío consiste en saber si algunos de esos ogros existen en la realidad.

Esto ha contribuido a la falta de progreso en el problema, pues los filósofos, en parte engañados por su propio tráfico de miedos y en parte valiéndose de esa urgencia engañosa para "motivar" la elaboración de sistemas y teorías metafísicos, se han fijado metas inalcanzables: la creación de imposibles talismanes filosóficos para evitar males inexistentes.

No pretendo insinuar que los filósofos avivaron *ex profeso* las llamas de la ansiedad o que explotaron astutamente la ansiedad con el fin de procurarse una motivación espuria para sus ejercicios metafísicos. Antes bien, los filósofos somos las víctimas y no los artífices de esas fantasías inducidas. Después de todo, somos los principales lectores de la literatura que, inocentemente, tiende a crear conceptos erróneos. Y, si somos cómplices de prolongar la vida de los errores, esto se debe, en parte, al deseo natural y universal de comprometernos en un proyecto cuya importancia pueda ser comprendida por quienes nos rodean. Si esto nos lleva a dramatizar un poco las cosas, a agravar ciertos contrastes y a marcar ciertos límites, sólo estamos haciendo lo que todos hacen en su propia línea de trabajo.

Nótese, por ejemplo, que una de mis premisas iniciales —que la gente se preocupa profundamente por el libre albedrío— ya sufrió una exageración en mis manos. Esto no implica que *todos* se preocupen por el libre albedrío como se preocupan por evitar el dolor o buscar el amor, por ejemplo. La mayoría de las personas —sin duda, más del 99%— está demasiado ocupada en mantenerse con vida y en defenderse de las circunstancias difíciles, de modo que no tiene ni tiempo ni ganas de dedicarse al problema del libre albedrío. Para la mayoría, la libertad metafísica, a diferencia de la libertad política, es un tema del cual no vale la pena ocuparse. Como dijo Dewey en cierta ocasión: "Aquello que los hombres valoraron y por lo cual lucharon en nombre de la libertad es diverso y complejo; pero, ciertamente, nunca fue la libertad metafísica" (Dewey, 1922, pág. 303).

Las personas, en su mayoría, no se preocupan por el libre albedrío. Pero nos resulta reconfortante (amable lector) creer que gracias a nuestro ocio y a nuestra vocación intelectual hemos indagado en sus circunstancias con mayor profundidad que ellas. Quizá sea cierto. No obstante, debemos ser cautos y no aceptar literalmente la intuición espontánea y compartida de que el problema del libre albedrío es uno de los *grandes problemas*. Constituimos un grupo selecto. Nótese, por ejemplo, que el problema del libre albedrío es una preocupación casi exclusivamente occidental. ¿Estamos acaso engañados? ¿Somos los únicos que creen en la importancia del libre albedrío? ¿Pensamos realmente que importa? Es decir, fuera de la sala de conferencias, de nuestras actividades profesionales o de nuestras disparatadas sesiones a medianoche, ¿el problema nos preocupa realmente? Como señaló Ryle, todos tenemos nuestros momentos fatalistas: "Sin embargo, aunque sabemos lo que significa acariciar la idea, aún no nos apasionamos por ella. No somos sus fanáticos seguidores ni sus fanáticos detractores" (Ryle, 1954, pág.

28). El fatalismo, según Ryle, no es una "tema candente", y lo mismo puede decirse del problema más amplio del libre albedrío. No obstante, es posible, por cierto, hacerlo aparecer como un tema candente.

Si el libre albedrío significa tanto para nosotros, debe ser porque no tenerlo sería terrible y porque puede haber razones para dudar de su existencia. ¿Qué es lo que, realmente, tememos? La falta de libre albedrío. ¿Y por qué? Quien teme la falta de libre albedrío debe tener alguna idea de *lo que significaría esa horrible condición*. En la literatura nos encontramos con una legión de analogías: no tener libre albedrío es como estar preso, hipnotizado, paralizado, o como ser una marioneta... (la lista continúa).

Pienso que estas analogías no son meros ejemplos útiles o meros recursos estilísticos sino que se hallan en la misma raíz del problema y constituyen un punto de apoyo para las discusiones filosóficas. Sin ellas, el problema del libre albedrío flotaría a la deriva o, en el mejor de los casos, se transformaría en un curioso problema destinado a entretener a metafísicos y traficantes de enigmas. Para una mayor comprensión, supongamos que un filósofo asegura haber resuelto el problema del libre albedrío; un lego podría aducir: "Muy bien, pero ¿su 'solución' alivia mis preocupaciones? Si no lo hace, no es la solución al problema del libre albedrío tal como me lo enseñaron". Por lo tanto, si seguimos la tradición, el libre albedrío es un problema que nos preocupa esencialmente. Los problemas acerca de la libertad que tienen un interés puramente esotérico no constituyen *el* problema del libre albedrío, por muy fascinantes que sean para algunos especialistas.

Más aun, los miedos no sólo proporcionan un punto de apoyo al problema del libre albedrío sino que también suministran el contenido y conforman la dinámica de la argumentación y de la investigación. El "clásico" y "tradicional" problema del libre albedrío es un engendro construido por los métodos tradicionales y las obsesiones de los filósofos. Este será uno de los temas de este libro.

Me propongo descubrir esos miedos, investigar el papel que desempeñan y, de ese modo, disipar algunas —no todas— preocupaciones y confusiones que ayudan a crear "el problema del libre albedrío". Dicho problema resultará ser una mezcla ilegítima y mal denominada de pánico autoinducido y de un planteo apresurado del problema; es decir, un falso pretexto para erigir sistemas innecesarios y chapucerías metafísicas.

Tenemos experiencia de cosas que son indiscutiblemente horribles y tememos que algunas de ellas constituyan nuestro destino; por eso, nos asusta la falta de libre albedrío. Conocemos muy bien esas circunstancias y poseemos buenas razones para aborrecerlas y

aterrorizarnos ante la perspectiva de que nos toquen en suerte, y es solamente por ello que nos preocupa el problema del libre albedrío.

Presentaré, a continuación, una lista de esos ogros y los analizaré sucintamente. Cada uno de ellos conduce una parte de la discusión tradicional sobre el libre albedrío. No es fácil erradicarlos por completo, pero al investigar los miedos en primer lugar, es posible que algunos se desvanezcan. Si (como decía mi madre) los miramos directamente a los ojos sin desviar un ápice la mirada y sin ponernos a pergeñar teorías, veremos que algunos de ellos son sólo productos de nuestra imaginación. Al tenerlos presentes desde el principio, nos será fácil distinguir sus sombras en los problemas que trataremos en los siguientes capítulos.

En su libro *The Concept of Mind* [*El concepto de lo mental*], Ryle intentó sacudirnos y avergonzarnos por un mal hábito de pensamiento adoptando la táctica de referirse a la concepción que estaba atacando "con deliberada grosería como 'el dogma del fantasma en la máquina' ".[1] En este primer capítulo, me expreso de manera similar —es decir, con la misma falta de respeto y caricaturizando un poco las cosas— cuando hablo de los ogros. En mi opinión, estas metáforas han hecho la mayor parte del trabajo detrás del escenario, atizando el problema del libre albedrío, y no merecen el menor respeto ni la influencia de que gozan. Así, mis caracterizaciones peyorativas tienen por objeto volvernos más sensibles a ellas y socavar su tradicional prestigio. Una vez inmovilizadas, las trataré con un espíritu más quirúrgico en los capítulos siguientes.

2. *Los hombres malos*

Los primeros ogros son los *hombres malos* —o las personas malas, si así lo desean— pues son concebidos como *agentes* que nos disputan el control del cuerpo, compiten con nosotros y tienen intereses opuestos o al menos independientes de los nuestros. Los filósofos utilizan con frecuencia a estos seres temibles como, diríamos, "desanimadores" (en oposición a "animadores") que se introdu-

[1] (Ryle, 1949, págs. 15-16.) La táctica no era a prueba de tontos. El neurofisiólogo sir John Eccles, en sus Conferencias Waynflete de 1952, en el Magdalen College de Oxford, presentó su moderna versión del interaccionismo cartesiano y cerró su exposición con lo que supuso eran palabras de cortesía para el profesor de filosofía: "Si uno utiliza la terminología de Ryle, el 'fantasma' maneja una 'máquina' que no tiene sogas, poleas, válvulas ni tuberías sino microscópicas configuraciones espacio-temporales de actividad en la red neuronal... Parecería ser, entonces, la clase de máquina que un 'fantasma' podría manejar" (Eccles, 1953, pág. 285).

cen en la escena cuando la ansiedad decae, cuando la urgencia del tema en discusión se vuelve dudosa. Como el embrollo se suma al embrollo, el lector comienza a bostezar y a impacientarse, pero es rápidamente galvanizado por una analogía que llama poderosamente su atención: "¡Pero eso sería como encontrarse de pronto en las garras de un...!".

El carcelero invisible. Las cárceles son horribles. Hay que evitar las cárceles. El que no lo entiende así no es uno de los *nuestros.* Ahora bien, si la cárcel es mala, ¿qué se le contrapone? Si uno no está preso, está libre (en un sentido importante). Y cada uno de nosotros puede reflexionar agradecido sobre la alegría de no estar en prisión. "¡Ajá!", dice el traficante de miedos. "¿Por qué está usted tan seguro de no ser un prisionero?" A veces es obvio que uno está en prisión, pero otras veces no lo es. Un carcelero astuto puede ocultar las rejas en los montantes de la ventana e instalar puertas falsas en los muros (si uno abre la puerta, se encuentra con una pared de ladrillos). Puede transcurrir un tiempo antes de que el prisionero se dé cuenta de que está en la cárcel.

¿Está usted *seguro* de no encontrarse en algún tipo de cárcel?[2] Se nos invita entonces a considerar una cadena de transformaciones que parte de cárceles obvias, pasa por cárceles no tan obvias (pero terribles) y termina en cárceles absolutamente invisibles e imposibles de localizar (pero ¿terribles?). Imaginemos un ciervo en el parque del Magdalen College. ¿Está preso? Sí, pero no demasiado, pues el cerco es muy amplio. Transportemos el ciervo a un espacio mayor, digamos el New Forest rodeado por una verja. ¿El ciervo seguiría preso? Me han contado que en el Estado de Maine los ciervos rara vez se alejan más de 5 millas de su lugar de nacimiento en toda su vida. Si se colocara un cerco más allá del límite normal del vagabundeo de un ciervo a lo largo de su vida, ¿se consideraría que el ciervo está preso? Quizá, pero nótese que el hecho de que *alguien* instale un cerco implica una gran diferencia. ¿Se siente uno preso en el planeta Tierra como Napoleón en Elba? Una cosa es nacer y vivir en la isla de Elba y otra ser llevado y retenido allí *por alguien.* Una cárcel sin carcelero no es una cárcel. Que sea o no una residencia indeseable dependerá de otros factores; en particular, de cómo coarta el estilo de vida de sus habitantes.)

El neurocirujano nefando. ¿Le agradaría que alguien lo atara con correas, insertara electrodos en su cerebro y controlara cada uno de sus actos y pensamientos apretando botones de la consola

[2] Berlin (1954, pág.68) dice que el determinismo "a pesar de sus cadenas cubiertas de flores, de su fachada de noble estoicismo y del esplendor y la inmensidad de su designio cósmico, representa el universo como una prisión".

"maestra"? Consideremos, por ejemplo, la clásica descripción de Fischer (1982, pág. 26): el ominoso Dr. Black, que altera el cerebro del pobre Jones "para poder controlar las actividades de Jones. Jones, por su parte ignora todo el asunto". En primer lugar, podemos preguntar —y siempre deberíamos hacerlo— por qué se introdujo este *agente* rival. ¿Por qué meter al Dr. Black en el asunto? ¿Dejaría de ser válido el ejemplo si, digamos, Jones tuviera un tumor cerebral que le produjera extraños comportamientos? La versión de Fischer es más terrible pues el control de las actividades de Jones ha sido usurpado por otro: el Dr. Black. Un tumor puede causar daños de toda índole en el cerebro de una persona y sería en verdad horrible padecer un tumor cerebral inhabilitante, pero tendría que ser un tumor sumamente astuto para poder *controlar* el cerebro de una persona.

Algunas de las variantes del neurocirujano nefando son el horrible hipnotizador y el titiritero prepotente. Todos hemos presenciado alguna vez exhibiciones de hipnotizadores y lo que resulta particularmente escalofriante es que, a diferencia del neurocirujano nefando, el hipnotizador no deja huellas de su influencia. Recordemos que se estipuló la ignorancia de Jones con respecto a la intervención del doctor Black (un punto importante que retomaremos en otros capítulos). Pero lo más insidioso es que los hipnotizadores profesionales exhiben a sus víctimas en público y las inducen a hacer el ridículo en presencia de personas que se encuentran en circunstancias más favorables. En esos casos, puede que le "ayude" imaginar las carcajadas del público mientras usted exhibe sus debilidades.

El titiritero prepotente es algo diferente, ya que controla sus dislocados *movimientos* a pesar de sus *esfuerzos y deseos*. Como el perro del estoico, usted luchará en vano en las garras del titiritero, pero al menos podrá tener la esperanza de revelar sus objeciones de conciencia al público, ora deslizando una mueca de disgusto, ora lloriqueando, un consuelo aparentemente inalcanzable para las víctimas del hipnotizador.

Nunca vimos una marioneta humana, pero conocemos la esclavitud y sabemos que es una de las condiciones más aborrecibles que se pueda imaginar. ¿Qué preferiría ser: el *zombie* del doctor Svengali o el lamentable títere humano? ¿Un esclavo o un prisionero? Se trata de destinos en cierto modo diferentes, aunque son todos detestables. Pero existen más villanos a quienes temer.

El niño cósmico cuyas muñecas somos nosotros. Nozick escribe: "Sin libre albedrío nos sentimos disminuidos, meros juguetes de fuerzas externas" (Nozick, 1981, pág. 291). ¡Qué indigno, no ser más que un juguete! Pero ¿cómo podríamos ser el juguete de una fuerza

impersonal? No puede haber juguetes sin jugadores; y los jugadores no son solamente agentes, sino agentes infantiles y juguetones.

Stanislaw Lem explora y explota el conocido supuesto filosófico de que podríamos ser meros juguetes en su cuento "The Seventh Sally or How Trurl's Own Perfection Led to No Good" (Lem, 1974). En la novela de Tom Robbins, *Even Cowgirls Get the Blues*, encontramos una deliciosa parodia de este clásico y filosófico cuento de terror.

> En la Navidad de aquel año, Julian le regaló a Sissy una aldea tirolesa en miniatura. Estaba construida con asombrosa maestría. Había una pequeña catedral cuyos vidrios de colores hacían una ensalada de frutas con la luz del Sol. Había una plaza y una cervecería. Los sábados por la noche, la cervecería se volvía sumamente ruidosa. Había una panadería que olía siempre a pan caliente y a *strudel*. Había un municipio y un departamento de policía con sus distintas secciones que contenían cantidades normales de expedientes y de corrupción. Había pequeños tiroleses en típicos trajes de cuero, cosidos en forma muy complicada, y, debajo de los pantalones, estaban los genitales hechos con la misma primorosa artesanía. Había tiendas de artículos para esquiar y muchas otras cosas interesantes, incluyendo un orfelinato. El orfelinato estaba diseñado para incendiarse por completo cada Nochebuena. Los huérfanos se arrojaban a la nieve con sus camisones en llamas. Terrible. Alrededor de la segunda semana de enero, aparecía un inspector del Cuerpo de Bomberos y hurgaba entre las ruinas, murmurando: "Si sólo me hubieran escuchado, esos niños aún estarían vivos". (Robbins, 1976, págs. 191-192.)

La maestría de este pasaje es notable. Nótese que la reiteración anual del drama del orfelinato —que repite la idea del eterno retorno de Nietzsche (todo lo que ha pasado volverá a pasar una y otra vez)— parece despojar a este pequeño mundo de todo significado real. Pero, ¿por qué precisamente la reiteración del lamento del inspector es lo que lo hace sonar tan vacío?

Quizá, si observamos más detenidamente lo que ello implica, descubriremos el artificio que hace "funcionar" este pasaje. ¿Los pequeños tiroleses reconstruyen el orfelinato o hay algún botón en la pequeña aldea que al oprimirlo vuelve a montarlo? ¿De dónde vienen los nuevos huérfanos? ¿Acaso los "muertos" regresan a la "vida"? Como veremos más adelante, un examen minucioso de tales fantasías revela a menudo que el verdadero trabajo lo hace un elemento tácito, estrictamente irrelevante para la tesis filosófica motivada, probablemente, por su invocación.

El maligno adivinador del pensamiento. Este agente es, en esencia, un rival que no provoca ni controla nuestros movimientos sino que los prevé y los frustra. Es inútil jugar con él a "piedra,

papel o tijeras", pues sabe exactamente en qué posición se encuentra uno, qué política se ha decidido seguir, cuál será el próximo movimiento y, en consecuencia, contraataca y gana (Hofstadter, 1982a). ¡Si tan sólo pudiéramos ocultarle nuestra mente! ¡Si pudiéramos hallar una estrategia que sus cálculos no pudieran predecir! Entonces no nos sentiríamos tan impotentes y vulnerables en el juego de la vida. Las predicciones tienen valor en la medida en que está en juego algo importante, cuando no son simples oraciones en tiempo futuro sino, más bien, apuestas que uno desea ganar y el oponente desea que se pierdan. En la vida real, competimos a menudo con otras personas e incluso con otros organismos (con las ratas y los mosquitos, por ejemplo). Pero, en el juego cósmico de la vida, ¿contra quién apostamos?

No puedo probar que ninguno de los miembros de esta galería de villanos existe realmente, como no puedo probar que el diablo o Santa Claus no existen. Pero estoy dispuesto a ponerme mi cara más seria y a asegurarle a quien necesite seguridad que no hay la más remota evidencia de que existan algunos de esos terribles agentes. En caso contrario, ¡pobre de nosotros! Un armario con un fantasma adentro es algo aterrorizante. Ahora bien, un armario que es como un armario con un fantasma adentro (pero sin el fantasma) no es, en absoluto, algo temible. De este modo, llegamos a un método empírico que puede sernos de gran utilidad: cuando usted vislumbre un hombre malo en algún ejemplo filosófico, compruebe si es este agente, seguramente ficticio, el que está haciendo todo el trabajo.

3. La avispa sphex *y otras preocupaciones*

El problema del libre albedrío se alimenta también de miedos que no tienen objetos personificados. La gente suele pensar que si el determinismo fuera verdadero, debería haber algo "mecánico" y, por lo tanto, lamentable, en nuestros procesos de deliberación. No seríamos agentes libres sino meros autómatas con una conducta similar a la de los insectos. Consideremos, por ejemplo, la avispa cavadora *sphex ichneumoneus*:

> Cuando llega el tiempo de poner los huevos, la avispa *sphex* construye un túmulo a tal efecto y parte en busca de un grillo. Luego de clavarle el aguijón a fin de paralizarlo, pero manteniéndolo vivo, lo arrastra al túmulo, pone los huevos junto a su cuerpo, cierra el túmulo, sale volando y no vuelve jamás. A su debido tiempo, los huevos se rompen y las larvas se alimentan del grillo que la avispa ha mantenido en un estado equivalente a un congelamiento profundo.

> Una tarea tan cuidadosamente organizada con vistas a un fin tiene, para la mente humana, un convincente sabor a lógica y a previsión (hasta que se examinan los detalles). Por ejemplo, la tarea de la avispa consiste en transportar el grillo paralizado hasta el túmulo, dejarlo en el umbral, entrar al túmulo para constatar que todo está en orden, salir nuevamente y sólo entonces arrastrar al grillo hacia el interior. Si se aleja al grillo unas pocas pulgadas del umbral cuando la avispa se encuentra en el interior, ésta, al salir, volverá a arrastrar al grillo hasta el umbral (pero nunca hacia el interior) y repetirá el procedimiento de entrar al túmulo para verificar si todo está en orden. Si nuevamente se aleja al grillo unas pocas pulgadas del umbral, mientras la avispa está adentro, volverá a arrastrarlo hasta el umbral y entrará otra vez para la inspección final. A la avispa no se le ocurre arrastrar al grillo directamente al interior. En cierta ocasión, este procedimiento se repitió cuarenta veces, siempre con el mismo resultado. (Wooldridge, 1963, pág. 82.)[3]

Hemos desenmascarado a la pobre avispa: no es un agente libre, sino que se encuentra a merced de la cruda causalidad física; sus estados y actividades dependen inexorablemente de factores del entorno que no puede controlar. En "Can Creativity Be Mechanized?", Hofstadter (1982b) denomina a esta desalentadora propiedad que tan vívidamente se manifiesta en la avispa *sphexicidad* [*sphexishness*]. El miedo a ser semejantes a la avispa *sphex* es una de las más poderosas fuerzas subterráneas en la literatura sobre el libre albedrío.

Gracias a Dios, somos mucho más inteligentes que la *sphex*; pero mientras reflexionamos agradecidos sobre este hecho, el traficante de miedos vuelve a la carga: "¿Por qué está tan seguro de no ser como la *sphex* (al menos en ciertos aspectos)?". ¿No se seguiría del determinismo materialista que los seres humanos, en tanto que organismos físicos, están como las avispas a merced de los estímulos y obstáculos ambientales que les caen encima? El biólogo, como un dios, desciende de las alturas, provoca un ínfimo desorden en el mundo de la avispa y nos revela su mecanicismo esencialmente irreflexivo. Una inteligencia superior que nos observara desde arriba ¿podría encontrar un truco semejante (aunque más elaborado) que nos desenmascarara? Y aun cuando pensemos que, hasta donde se sabe, no existen allá arriba tales agentes superiores decididos a frustrar nuestras vidas, la sola posibilidad de ser tan imperfectos y vulnerables nos perturba notablemente.

[3] El mismo pasaje aparece también en Wooldridge, 1968. Para una descripción de otros aspectos filosóficamente interesantes de la conducta de la avispa, véase Dawkins, 1980: "Good Strategy or Evolutionarily Stable Strategy?".

Nótese el paralelismo entre el carcelero invisible y el temor a ser como la avispa *sphex*. Se empieza con un ejemplo simple y claro de algo horrible (estar literalmente presos o ser *exactamente* como la avispa) y luego, permitiendo que eso horrible penetre en nuestras profundidades, aceptamos que, si bien en nuestro caso las cosas son mucho más complejas —y hasta difíciles de imaginar— , son similares en un sentido importante. Así, gracias a la cadena de similitudes, nuestro propio caso hereda aquella cualidad horrible. Pero, ¿son realmente válidas esas semejanzas?

Quiero llamar la atención sobre una peligrosa práctica filosófica que examinaré exhaustivamente en este libro: la deliberada simplificación del papel que cumple la imaginación del filósofo. En filosofía, existe una estrategia muy popular que consiste en construir un cierto tipo de experimento mental que he denominado *generador de intuición* (Dennett, 1980 y Hofstadter y Dennett, 1981). Estos experimentos mentales, a diferencia de los de Galileo o los de Einstein, no presentan argumentos estrictos que prueban conclusiones a partir de premisas. Antes bien, su objetivo es brindar al lector un conjunto de reflexiones imaginativas que desemboca, finalmente, no en una conclusión formal sino en un dictado de la "intuición". Los generadores de intuición están astutamente diseñados para centrar la atención del lector en los aspectos "importantes" e impedir que el lector se empantane en detalles difíciles de seguir. En principio, no hay nada de malo en este procedimiento. En efecto, una de las misiones más altas de la filosofía es la de ayudar a la gente a ver el bosque y no solamente los árboles. Pero, con frecuencia, se abusa de los generadores de intuición, aunque raras veces en forma deliberada.

Quizás el abuso más frecuente se comete al derivar un resultado —un juicio intuitivo sincero— a partir de la misma simplicidad del caso imaginado y no a partir del contenido real del ejemplo descrito con tanta claridad y sencillez. Tal vez el destino de la avispa nos parece tan terrible no porque sus acciones y "decisiones" sean *causadas* sino precisamente por el hecho de ser causadas de una manera tan *simple*. Si esto es efectivamente así, el reconocimiento de la diferencia entre el objeto de nuestro generador de intuición y nosotros mismos —nuestra complejidad— puede impedir que heredemos la temible cualidad que observamos en el caso simple. Quizá deberíamos reírnos en lugar de estremecernos. Quizás este generador de intuición es como la serpiente que se traga la cola y no se detiene hasta devorarse por completo.

Pero sólo un análisis detallado podrá dirimir la cuestión. ¿Somos como las avispas *sphex* ? ¿Hasta qué punto? Conocemos, por cierto, algunas personas que lo son: los dementes, los retardados,

los que han sufrido daños cerebrales. Whitaker (1976) describe el caso de una mujer que sufría daños cerebrales y que ya no comprendía ningún lenguaje, pero que era capaz de repetir como un loro todo lo que se le decía (¡excepto los errores gramaticales, que nunca dejaba de corregir!). Muchos experimentos perturbadores realizados por psicólogos revelan algo de esa dimensión en la que nos comportamos como la *sphex* : el clásico cuento de terror de Milgram sobre los torturadores obedientes (Milgram, 1974), los experimentos sobre la irracionalidad humana efectuados por Kahneman, Tversky y otros (Kahneman, Slovic y Tversky, 1982)[4] y, por supuesto, las famosas —aunque no confirmadas oficialmente— anécdotas de los estudiantes que usaban las técnicas de condicionamiento de Skinner para lograr que sus profesores de psicología se rascaran la oreja mientras dictaban sus clases (Brewer, 1974). En qué medida somos como la avispa *sphex* será el tema principal del segundo capítulo.

El yo que desaparece. Otro elemento oculto en el cuento de la avispa es esa sensación fantasmagórica que experimentamos con frecuencia cuando observamos o estudiamos a los insectos y otros animales inferiores: ¡toda esa actividad bulliciosa y sin embargo *no hay nadie en casa* ! Estamos mirando un mundo que al parecer fue planeado con inteligencia y luego abandonado por quien lo planeó. Las hormigas, las abejas e incluso los peces y las aves sólo "padecen sus movimientos"; no comprenden ni valoran lo que son capaces de hacer y no hay, en los alrededores, otros yoes con capacidad de comprensión. Este es el miedo al increíble yo que desaparece.

Además, conocemos casos claramente intermedios entre los insectos y los peces y nosotros. Por ejemplo, los que sufren de insania y de daños cerebrales pueden ser apropiadamente descritos como carentes de yo, como seres vivos y animados pero sin *animae*. Cuando observamos "con excesiva atención" nuestras propias actividades mentales, se produce también, a menudo, una desaparición del yo. En cierta ocasión, refiriéndose a sus ideas musicales, Mozart dijo: "¿De dónde vienen y cómo vienen? No lo sé y *no tengo nada que ver con eso*" (la bastardilla es mía).[5] Si el determinismo es verda-

[4] Milgram comprobó que un marco lo suficientemente "institucional" puede provocar conductas terriblemente dóciles en personas en apariencia normales. Varios sujetos fueron empleados como "asistentes de investigación" y se les pidió que aplicaran severos *electroshocks* a otros "sujetos" (que, en realidad, eran actores) como parte de un programa de entrenamiento. Kahneman y Tversky demostraron en muchos de sus experimentos que incluso investigadores avezados son a menudo engañados con las falacias elementales del razonamiento inductivo, tales como la falacia del jugador (si al tirar una moneda ha salido "cara" con mucha frecuencia, es probable que salga "cruz" en el próximo tiro). Para una reseña y una discusión filosófica de estos experimentos, véanse Cohen, 1981 y Kyburg, 1983.

[5] Véase mi discusión sobre Mozart, Poincaré y la creatividad en Dennett, 1975.

dero, entonces, no hay libertad de acción para nuestros yoes ni tarea alguna que puedan realizar. ¿Podemos encontrar a nuestros yoes o la ciencia está a punto de demostrar que son meras ilusiones, semejantes a la ilusión de que la avispa es (o tiene) un yo?

La ciencia nos conduce al interior de las cosas, y la visión interna y detallada del cerebro que nos ofrece no revelará, probablemente, ninguna versión reconocible de lo que Descartes llamaba *res cogitans* o cosa pensante que tan bien conocemos "por introspección". Pero, si perdemos de vista nuestros yoes en favor de la objetividad científica, ¿qué ocurrirá con el amor y la gratitud (y con el odio y el resentimiento)?

> Pero, ¿qué haré si ni siquiera siento resentimiento?... A causa de las malditas leyes de la física, mi ira está condenada a la descomposición química. Cuando la observas, su objeto se desvanece en el aire, sus razones se evaporan, el ofensor no está en ninguna parte, la afrenta deja de ser una ofensa y se convierte en destino, en algo semejante a un dolor de muelas del que nadie es culpable. (Dostoievski, *Notes from Underground [Memorias del subsuelo]*, citado por Bennett, 1980).

El secreto terrible. Con frecuencia la ciencia parece estar a punto de decirnos demasiadas cosas, de abrir la caja de Pandora y revelarnos algún secreto terrible. Tan pronto como lo conozcamos, nos paralizará. Nos paralizará al destruir alguna ilusión absolutamente necesaria para nuestras vidas como agentes. Nuestra propia racionalidad nos destruirá, pues una vez que hayamos contemplado la verdad, ya no podremos continuar engañándonos.

Es bastante sencillo comprobar de qué manera el secreto terrible provoca la parálisis, pues es análoga a algo que ocurre frecuentemente y que es realmente paralizante e incluso patético. Supongamos, por ejemplo, que estamos en Oxford deliberando acerca de los méritos relativos de los restaurantes londinenses y tratando de decidir a cuál de ellos concurriremos esta noche. Se nos informa, entonces, que los trenes no circulan, que los restaurantes no abrirán sus puertas durante todo el día, que estamos encerrados en nuestro cuarto en Oxford o, simplemente, que es demasiado tarde para viajar a Londres. Así, nos damos cuenta de que nuestras deliberaciones han sido completamente inútiles. No había oportunidad real para actuar ni alternativas reales para elegir. Ahora bien, si la ciencia nos demostrara que no hay, en verdad, oportunidades, ¿no nos conduciría esto —o debería conducirnos— a abandonar todo tipo de deliberación?

Como dice Tolstoi en la última línea de *La guerra y la paz*: "Es necesario renunciar a una libertad que no existe y reconocer una dependencia que no sentimos". Pero esto es terrible, pues ¿no nos

induciría a una apatía y a una resignación verdaderamente perniciosas y destructivas? Pensemos, por ejemplo, en la obscena resignación de quienes consideran inevitable la guerra nuclear y, en consecuencia, opinan que no vale la pena tratar de impedirla. ¿No deberíamos deplorar las declaraciones que alientan este tipo de actitud (aun si son verdaderas, o mejor, especialmente si son verdaderas)?

¿En qué consiste el secreto terrible? Quizás, en la realidad del determinismo (¡o en la realidad del indeterminismo!). Sea como fuere, es necesario que el secreto terrible quede al descubierto, ya que implica que la libertad es una ilusión. Nótese que el miedo no está referido a la verdad o falsedad de cierta proposición, sino al hecho de que, verdadera o falsa, pueda ser creída. Por otra parte, si el determinismo es verdadero en el presente, también lo fue en el pasado. Hay quienes padecieron una vida horrible, pero hay otros, en cambio, que gozaron de una vida digna de ser vivida, aun en un mundo determinista. La ciencia moderna sólo está descubriendo que el determinismo es verdadero, no lo está *haciendo verdadero*, de modo que las cosas no van a empeorar, a menos que sea la creencia en el determinismo y no el determinismo en sí lo que provoque la catástrofe.[6]

Cabe preguntarnos si el descubrimiento del determinismo no sólo puede arruinar nuestras vidas sino también revelar, retrospectivamente, que aquella buena vida del pasado no era tan buena como les parecía a quienes la vivieron. Esta preocupación está presente en algunas de las imágenes más frecuentes de la literatura filosófica. Anscombe (1957, pág. 6) nos habla de una conferencia en la cual Wittgenstein invitó al público a pensar en las hojas cayendo en otoño y diciéndose a sí mismas: "Ahora iré por aquí... ahora iré por allá". Hobbes inventó una fantasía similar.

> Un trompo de madera que los niños hacen girar tirando de una cuerda y se desplaza girando, ora hacia una pared, ora hacia la otra, ora golpeando a las personas en las piernas, si fuera sensible a sus propios movimientos, pensaría que actúa según su voluntad, a menos que sintiese la cuerda que lo puso en movimiento. ¿Y un hombre más sabio cuando corre a un lugar por un beneficio, luego a otro por un negocio y perturba al mundo escribiendo errores y exigiendo respuestas porque piensa que sólo lo mueve su voluntad y no ve cuáles son las cuerdas que la mueven? (Hobbes, Molesworth, comp., 1841, vol. V, pág. 55.)

[6] Véase, por ejemplo, Strawson, 1962: "¿Qué efecto tendría, o debería tener, la aceptación de una tesis general del determinismo sobre las actitudes reactivas [nuestras actitudes normales de participación en relaciones interpersonales, tales como la gratitud o el resentimiento]?".

Algunas ilusiones son casi irresistibles. El golfista, luego del golpe, observa cómo la pelota zigzaguea lentamente en dirección al hoyo. Se tuerce, se dobla y se inclina como si quisiera alterar el rumbo de la pelota, como si sus contorsiones pudiesen modificar la trayectoria. Pero, obviamente, es demasiado tarde. Existe un término para denominar a estos movimientos grotescos: "inglés corporal" [*body English*].[7] El inglés corporal es siempre inútil, a veces cómico, a veces patético y a menudo irresistible. La ciencia amenaza con mostrarnos que todos nuestros esfuerzos son sólo mero inglés corporal. ¿No sería espantoso si toda nuestra gimnasia mental, nuestras deliberaciones, luchas, resoluciones y esfuerzos no fueran más que mero inglés corporal? Lo serían, si no pudiesen modificar los resultados de los acontecimientos que nos importan. Este ogro aparece en todas las discusiones sobre el fatalismo, pero no es éste su único coto de caza. Por el momento, puede ser útil comparar el inglés corporal con algo bastante análogo a fin de librarnos del miedo.

Consideremos, ahora, el *follow-through*.[8] El profesor de golf aconseja al golfista *mantener la cabeza baja* hasta haber completado el *swing*. ¿Es éste un buen consejo? La pelota abandona la cabeza del palo en la mitad del *swing* y, una vez comenzada su trayectoria, nada de lo que suceda en el *tee* podrá modificarla. ¿Prestar atención a los detalles del *swing* luego de que la pelota abandona la cabeza del palo no sería demasiado inglés corporal? No necesariamente, pues, quizá, la única forma de lograr que suceda lo correcto hasta el momento del impacto sea mirar hacia adelante y fijarse una meta, confiando en que los propios esfuerzos para alcanzar esa meta producirán movimientos corporales que atravesarán el espacio justo a la velocidad justa. Sería tonto de nuestra parte desatender el consejo del profesor, basándonos en el argumento antes mencionado de que no se gana nada siguiendo el consejo. Podría ganarse todo. A veces, la única manera de obtener lo que realmente deseamos es hacer algo más. (Ampliaremos estas observaciones en los capítulos 5 y 7.)

Por lo tanto, estoy dispuesto a combatir el fuego con el fuego. El traficante de miedos invoca la imagen familiar del golfista y sus grotescos movimientos corporales y logra que desplacemos nuestros estremecimientos de turbación o de ansiedad al reino metafísico del libre albedrío. Yo contraataco con la imagen familiar del

[7] Descubrí en Oxford que el término es un norteamericanismo ampliamente desconocido en Inglaterra, lo mismo que su pariente *"putting English" on a ball* [darle efecto a la pelota].

[8] Véase la breve discusión sobre la terminación del *swing* en Nozick, 1981, pág. 311.

follow-through y pregunto por qué no llevamos también la moraleja de esta imagen al elevado terreno metafísico. Pero tendría que haber un modo de proceder más adecuado. La práctica filosófica tradicional consiste en atravesar rápidamente las analogías hasta llegar a una conclusión que luego se convierte en el punto de partida para la argumentación y la construcción minuciosa de teorías. Por ejemplo, se considera "obvio" que el libre albedrío que todos deseamos es tal que uno tiene libre albedrío sólo si "podría haber actuado de otra manera", y luego se invierte una enorme cantidad de energía y de paciencia tratando de encontrar las condiciones necesarias y suficientes para este tipo de facultad o circunstancia.

Así, se origina en algunas personas el curioso deseo de que, para cualquier acto, sea tan verdadero que, si se repitiera *exactamente el mismo* estado de cosas físico, podría realizarse un acto diferente. Se ha gastado mucho ingenio tratando de elucidar a qué equivale esta tesis y qué probabilidad tiene de ser verdadera, pero, sorprendentemente, se ha prestado poca atención a la cuestión de por qué nos importa tanto esta posibilidad metafísica (además del hecho de que, si no fuera verdadera, ¡ay!, eso equivaldría a estar preso, paralizado, hipnotizado, a ser como una avispa, una marioneta o un juguete). Las alusiones a la alternativa terrible a veces son examinadas tan rápidamente que se pasa por alto una incoherencia tan obvia —incoherencia que no resistiría la cuidadosa atención que los filósofos dedican a la construcción correcta de teorías—.

4. Resumen

Hasta el momento, no he tratado de probar nada acerca del libre albedrío sino que me he limitado a rodear el tema, a esbozar el modo en que enfocamos el problema y a llamar la atención sobre ciertos aspectos curiosos de la materia prima (una veta en el mármol o una forma tentadora). También me ha interesado subrayar de qué manera el problema del libre albedrío podría ser, en buena parte, un producto artificioso de los métodos que se utilizan comúnmente para estudiarlo. Este primer despertar de la conciencia nos será útil para evitar los caminos trillados mientras recorremos su dominio tradicional. Nos ocuparemos de prácticamente todos los temas y argumentos tradicionales en la literatura sobre el libre albedrío, pero mi método será *demorarnos allí donde los demás se apresuran*, investigando a fondo las analogías en lugar de arrojarnos de cabeza a la construcción y refutación de teorías.

La represión profesional, vigorosamente institucionalizada, nos ha impedido apreciar la importancia y el poder de los generadores de intuición en el desarrollo de la filosofía. La historia de la

filosofía nos demuestra que los generadores de intuición han sido los motores fundamentales en todas las áreas y no solamente en el terreno del libre albedrío. Pensemos en la caverna de Platón, en Menón enseñando geometría al pequeño esclavo, en el estado de naturaleza de Hobbes, en el genio maligno de Descartes o, más recientemente, en los lingüistas de Quine (1960) tratando de traducir *"gavagai"*, en el enigma del *"grue-bleen"* * de Goodman (1965), en la posición original de Rawls (1971), en la pregunta de Farrell acerca de cómo sería ser un murciélago (1950),[9] para no mencionar la famosa Tierra gemela de Putnam (1975) y el aún más famoso cuarto chino de Searle (1980).

Se podría decir que estos generadores de intuición son la melodía perdurable de la filosofía; su poder de duración les asegura un lugar en la memoria de nuestros jóvenes estudiantes, que los recordarán vívidamente y con exactitud muchos años después de haber olvidado el intrincado contrapunto de la argumentación y del análisis. Un buen generador de intuición es más sólido que cualquiera de sus variantes. ¿Cuántas versiones se han sucedido a partir de la posición original de Rawls o de la Tierra gemela de Putnam?

Los generadores de intuición son instrumentos pedagógicos poderosos. La mayoría de los filósofos está de acuerdo en que el experimento mental de Descartes, el *"cogito ergo sum"*, es sospechoso desde un punto de vista lógico, si no totalmente incorrecto. Ha inspirado docenas de reinterpretaciones y defensas que los profesores de filosofía preferirían omitir, pero jamás se les cruzaría por la mente sacar de sus programas de estudio la dramática idea de Descartes. Los grandes generadores de intuición pueden, incluso, extraviarnos en lugar de instruirnos. Cuando enseñamos Descartes, no afirmamos que su experimento mental revela o conduce a la verdad acerca del conocimiento. En efecto, solemos culpar a Descartes y a su atractivo generador de intuición por haber arrastrado a los filósofos, durante tres siglos, a una lucha salvaje. En el mejor de los casos, le estamos agradecidos como podríamos estarlo con alguien

* Juego de palabras entre *green* (verde) y *blue* (azul). [T.]

[9] En efecto, fue Farrell quien, en 1950, planteó (y respondió) por primera vez la pregunta filosófica que luego se hizo famosa en el clásico artículo de Nagel (1974) "What is it Like to be a Bat?" (reeditado, con comentarios, en Hofstadter y Dennett, 1981). Farrell analiza esta pregunta en su artículo siempre vigente "Experience", publicado primero en *Mind* y reeditado en la famosa antología de Chappell (1962). Llamo la atención sobre la prioridad de Farrell no para molestar a Nagel —todos somos proclives a inventar lo que admiramos— sino para subrayar la solidez de un buen generador de intuición, semejante a la solidez de una melodía hermosa e inolvidable.

que nos dio direcciones equivocadas que, sin embargo, nos condujeron a un fascinante contratiempo del cual aprendimos mucho.

El papel fundamental que desempeñan los generadores de intuición en filosofía demuestra que ésta no es —ni puede razonablemente aspirar a ser— una ciencia. Sin los generadores de intuición, la filosofía puede, en ocasiones, purificar y regimentar con éxito un área conceptual para que la ciencia se haga cargo de ella, pero esto no constituye un auténtico triunfo filosófico. Con los generadores de intuición, la filosofía no es una ciencia, pero, en su estilo informal, es una valiosa —y a veces necesaria— compañera de la ciencia. A los filósofos no debería molestarles el hecho de que los generadores de intuición realicen la mayor parte de la labor filosófica perdurable (para mejor o para peor). Si la filosofía sirve para ensanchar nuestra visión de lo posible y para romper malos hábitos de pensamiento, los generadores de intuición se convierten, entonces, en la herramienta ideal entre todos los avíos del filósofo. Como dijo Wittgenstein: "La filosofía es una batalla contra el hechizo de nuestra inteligencia por medio del lenguaje" (Wittgenstein, 1953, secc. 109). Para una tarea semejante, el ordenamiento reglamentado de la argumentación rigurosa es, raras veces, algo más que una póliza de seguro, un cheque para solventar el libre tráfico de intuiciones que ha sentado las bases de una visión.

En los siguientes capítulos, sostendré que el problema del libre albedrío *es* la familia de ansiedades que hemos esbozado aquí someramente. Mi método consistirá en analizar esas ansiedades, y las analogías y los generadores de intuición que las alimentan, a fin de descubrir qué amenazas reales a nuestra autoestima y a nuestras aspiraciones se ciernen en el horizonte y qué problemas filosóficos de auténtico interés quedarían por resolver.

En el capítulo dos, me ocuparé de cuestiones referidas a nuestro *status* biológico en tanto que animales racionales y examinaré las razones del miedo a ser semejantes a la avispa *sphex*.

En el capítulo tres, examinaré los conceptos de *control* y de *autocontrol*, dos conceptos de suma importancia para el libre albedrío y el determinismo que, hasta donde yo sé, nunca fueron analizados exhaustivamente por los filósofos. Las preguntas fundamentales son: ¿de qué manera una cosa controla a otra (o a sí misma)? y ¿qué tipo de cosas pueden ejercer control? (Aquí es donde los Hombres Malos continúan interfiriendo.)

En el capítulo cuatro, retomo el concepto de *yo* o *agente* y trataré de ver cómo se puede evitar su desaparición ante la arremetida de la ciencia.

En el capítulo cinco, analizaré la afirmación de Kant de que debemos "actuar bajo la idea de la libertad" cada vez que actuamos.

¿De qué manera tenemos que concebir el futuro y nuestras facultades para poder dividir el mundo en cosas que "dependen de nosotros" y en cosas que "no dependen de nosotros"? ¿La libertad de acción que creemos tener cuando deliberamos existe realmente? Alfred Prufrock pregunta:

> ¿Me atrevo
> a perturbar el universo?
> En un minuto hay tiempo
> para decisiones y revisiones que un minuto trastrocará.
> ("The Love Song of J. Alfred Prufrock", T. S. Eliot)

Pero, ¿cómo podría algo ser imputable de perturbar el universo? La ciencia parece necesitar la idea (véase *Disturbing the Universe*, Dyson, 1979) y, al mismo tiempo, la desecha. ¿Es posible armonizar ambas actitudes?

En el capítulo seis, examinaré el significado del verbo "poder" y de la frase "podría haber actuado de otra manera". Como dijo admirablemente Austin: "En filosofía, 'poder' es aquello que tan a menudo descubrimos —cuando ya estábamos seguros de haber resuelto el problema— riéndose sarcásticamente de nosotros como la rana en el fondo del jarro de cerveza" (Austin, 1961, pág. 179). Es cierto, pero primero tenemos que vaciar el jarro; entonces nos encontraremos con una rana mucho más fácil de manejar.

En el capítulo siete, pregunto, en primer lugar, por qué deseamos el libre albedrío y alego que se trata de un deseo sabio, dadas las implicaciones de nuestra necesaria imperfección en cuanto agentes. En este capítulo, las preguntas son de índole práctica y personal. ¿Nos engañamos a nosotros mismos —o somos engañados por la sociedad— cuando insistimos en ser *considerados responsables*? ¿Cuándo y por qué disminuye nuestra responsabilidad? Cuando actuamos de manera equivocada, ¿somos verdaderamente culpables?

Mis conclusiones no son revolucionarias ni pesimistas; son sólo moderadamente revisionistas: lo que dice el sentido común acerca de nuestro lugar en el universo es aproximadamente cierto. Tenemos libre albedrío. Podemos tener libre albedrío y también ciencia. Estas conclusiones no agotan el tema del libre albedrío y cabe esperar futuros desafíos a nuestra serenidad. Pero este resumen brindará algún consejo sobre cómo enfrentarlos.

2

Hacer la razón práctica

1. ¿De dónde provienen las razones?

Un tema recurrente y aún controvertido en la literatura sobre el libre albedrío es la idea de que la verdadera libertad de la voluntad consiste (o debería consistir) en una completa sumisión de la voluntad a los *dictados de la razón*. Kant es, sin duda, el filósofo que hizo de esta idea la piedra angular de su teoría, pero no fue el primero que reconoció en la perfecta racionalidad el ideal con el cual medir todas las imperfecciones y ataduras de la voluntad o del obrar. Spinoza, en "Sobre la servidumbre humana", afirma que:

> Obrar en obediencia absoluta a la virtud es, en nosotros, lo mismo que obrar, vivir o preservar el propio ser (estos tres términos tienen el mismo significado) de acuerdo con los dictados de la razón, sobre la base de la búsqueda de la propia utilidad. (*Etica*, IV, prop. XXIV)

Descartes se expresa de manera similar en la "Cuarta Meditación"[1] y el doctor Bramhall, obispo de Derry, pronuncia una suerte de himno sobre el tema en su debate con Hobbes.

[1] "Pues para que yo sea libre, no es necesario que me mantenga indiferente en elegir uno u otro de dos contrarios; antes bien, cuanto más me inclino hacia uno, sea porque veo claramente que contiene (la preponderancia de) la bondad y la verdad, sea porque Dios guía así mis pensamientos, tanto más libremente lo elijo (y abrazo). Y, ciertamente, la gracia divina y el conocimiento natural, lejos de disminuir mi libertad, más bien la aumentan y la fortalecen. Más aun, esa indiferencia que siento, cuando no soy movido hacia un lado más que hacia otro por (el peso de) alguna razón, es el grado más bajo de la libertad, y más bien manifiesta un defecto en el conocimiento que una perfección en la voluntad; pues si yo conociera siempre claramente lo que es verdadero y bueno, jamás necesitaría deliberar acerca de qué juicio y qué elección debería hacer, y de ese modo yo sería enteramente libre, sin ser jamás indiferente." Véase también su carta a Mersenne (*Oeuvres de Descartes*, 1897-1913, vol. III, págs. 381-382):"Me muevo más libremente hacia un objeto en proporción con el número de razones que me compelen; pues es cierto que mi voluntad es movida con más facilidad y espontaneidad".

> La razón es la raíz, la fuente, el origen de la verdadera libertad, que juzga y manifiesta a la voluntad si esto o aquello es conveniente, si esto o aquello es más conveniente. (Hobbes, *Works*, vol. V, pág. 40)

Locke, cuya concepción no es "kantiana", sustenta, sin embargo, el ideal de vivir según los dictados de la razón y lo hace con una retórica tan efectiva como la de Kant:

> Si desprenderse de la guía de la razón y desear la suspensión del examen y del juicio que nos impiden elegir o hacer lo peor es libertad, verdadera libertad, los locos y los tontos son los únicos hombres libres; aun así, pienso, nadie elegiría ser loco en favor de semejante libertad, excepto aquel que ya está loco. (*Ensayo sobre el entendimiento humano*, II, XXI, secc. 51)

La idea de libertad como perfecta racionalidad aparece en la literatura con tanta frecuencia que debe de encerrar algo profundamente cierto, pero se la expresa con tanta vehemencia que debe de contener también un error, algo no intuitivo e inquietante que crea la resistencia contra la cual éstos y otros autores proclaman sus ideas. En el capítulo tres, voy a explorar las fuentes de este escepticismo con respecto al ideal de racionalidad (o, más exactamente, de la falta de apego al mismo), pero ahora quiero examinar su poder de seducción y mostrar cómo podemos aspirar a esa racionalidad de una manera realista.

¿Cómo logró la razón hacerse un sitio en un universo mecánico y material? En el principio no había razones, sólo causas. Nada tenía un propósito o una función; no existía teleología en el mundo. La explicación es sencilla: no había nada que tuviera intereses. Pero, después de milenios, aparecieron los simples replicantes [2] y, aunque *ellos* no tenían la menor idea de sus intereses (o, mejor dicho, no tenían ningún interés), nosotros, mirándolos retrospectivamente desde nuestro punto de vista "divino", podemos atribuirles, sin arbitrariedad, ciertos intereses originados a partir de su definido "interés" en la autorréplica. Es decir, quizá carecía de importancia y a nadie le interesaba si lograban o no replicarse (aunque podemos estar agradecidos de que lo hicieran); pero, al menos, podemos atribuirles intereses condicionalmente. *Si* estos simples replicantes deben sobrevivir y replicarse, resistiendo así la creciente entropía, su entorno tendrá que reunir ciertas condiciones: las condiciones que conducen a la replicación deben estar presentes o ser, al menos, frecuentes.

[2] Véase en Dawkins 1976, una lúcida y elegante versión de la historia. Para una perspectiva diferente, véase Monod 1976, cap. I, "Of Strange Objects", sobre el nacimiento de lo que Monod llama la "telenomía".

Dicho de manera más antropomórfica, si estos simples replicantes desean continuar replicándose, tienen que luchar por varias cosas, esto es, evitar las cosas "malas" y buscar las "buenas". Para decirlo dramáticamente, si imagináramos que somos los guardianes de sus intereses, podríamos ver claramente los pasos a dar y la asistencia y el asesoramiento que sería necesario brindarles.

Esto no es decir mucho, pues también es cierto que si nos imaginamos encariñados con alguna formación rocosa particularmente bella, vomitada hace millones de años por alguna erupción volcánica, podemos imaginarnos rápidamente los pasos que habría que seguir para preservarla (protegerla de la erosión, evitar que el sedimento la sepulte, que la destruyan futuras erupciones volcánicas, etcétera).

¿Cuál es la diferencia? ¿Cómo llegaron a cobrar vida propia los intereses de los replicantes? Simplemente así: los replicantes se fueron convirtiendo poco a poco en guardianes de sus propios intereses. En efecto, la capacidad de autorréplica dependía de ello. A diferencia de la escultura volcánica, no eran tan extremadamente desvalidos ni dependientes de la solicitud ajena; hasta cierto punto, podían cuidar de sí mismos. El día en que el universo tuvo entidades capaces de tomar algunas medidas rudimentarias para defender sus propios intereses fue el día en que nacieron los intereses. Las mismas tendencias de los organismos a preservar distintos estados (sus variedades de *homeostasis*) ayudaron a ajustar la definición de sus intereses. Sólo ciertos tipos de homeostasis tendieron a la autoconservación a largo plazo; éstos fueron replicados y, en consecuencia, persistieron y, en consecuencia, *dieron una nueva definición* del interés "primordial" por la autoconservación y la autorréplica. Así, si el mantenimiento de la temperatura corporal desempeñara un papel importante en la autoconservación de los miembros de una especie, perdurarían los sistemas de control del mantenimiento de la temperatura corporal que evolucionaran. Por lo tanto, el catálogo de intereses de esa especie llegaría a incluir el mantenimiento de un cierto grado de temperatura corporal.

Los temas fundamentales de esta historia han sido mencionados muchas veces. Buscar alimento, protegerse de los predadores, localizar al compañero, aparearse y mantener la salud (curarse a sí mismos, evitar el trauma, conservar la energía, etc.) son los principales objetivos secundarios de los replicantes. Dichos objetivos, en interacción con las circunstancias particulares de cada especie, generan, a su vez, objetivos secundarios instrumentales: detectar olores, cavar agujeros, desplazarse, reconocer configuraciones, sentir dolor, cortejar a la pareja, etcétera.

Así, cuando llegamos a una criatura tan evolucionada como nuestra *sphex*, podemos asignarle un catálogo de intereses bastante elaborado. Si no fuera así, si no tuviéramos una visión clara y detallada de cuáles son sus intereses, su actuación en el extraño experimento descrito en el capítulo uno no nos parecería tan patética. La actuación de la avispa resulta ser muy pobre si la juzgamos según sus propios intereses. Si fuéramos sus guardianes, podríamos aconsejarle mejores cursos de acción (¡si tan sólo pudiera seguir nuestro consejo!).[3] Al parecer, es un guardián muy imperfecto de sus intereses,[4] y podemos decir por qué: ignora cuáles son sus intereses e incluso si los tiene. ¡Uno se siente tentado a preguntarse si acaso posee alguna noción de su propia existencia!

No obstante, la supuesta ignorancia de sus intereses no le impide "tenerlos", del mismo modo que una persona en estado de coma no deja de tener intereses por el hecho de perder la conciencia de ellos o de ser incapaz de defenderlos. (Cuando a una persona tan desafortunada se le asigna un guardián, significa que hay algo que guardar, además del cuerpo con vida; un guardián no es un mero curador. Pero me abstendré de una digresión que nos llevaría a considerar temas tales como la supervivencia de los intereses de una persona después de su muerte.)

Cuando entra en escena una entidad cuyo comportamiento retarda, aunque sea en forma rudimentaria, su propia disolución y descomposición, esa entidad lleva al mundo su "bien". Es decir, crea un punto de vista desde el cual los acontecimientos del mundo se pueden dividir, a grandes rasgos, en favorables, desfavorables y neutros. Y sus propias tendencias innatas a buscar los primeros, eludir los segundos e ignorar los terceros contribuyen esencialmente a la definición de las tres clases. Al tiempo que la criatura comienza a tener intereses, el mundo y sus acontecimientos comienzan a crear *razones* que los justifican. Así, la pobre avispa tiene una razón para ocultarse del biólogo y una razón para alimentar a su progenie, pero fue diseñada para actuar (y con bastante eficacia) conforme a la segunda y, aparentemente, no tiene conciencia ni poder con respecto a la primera. Esto es natural. La evolución lleva tiempo e implica una presión considerable sobre nuestros ancestros (más precisamente, sobre los contemporáneos menos afortunados de nuestros ancestros). Es fácil ver por qué no hay nada en la herencia

[3] Quizá la subestimemos. Acerca de las trampas de un veredicto prematuro sobre la estupidez de la avispa, véase Dawkins 1982, págs. 48-50. Véase también Dennett 1983b.

[4] Pero quizá sea un buen guardián. Los costos de un cerebro más desarrollado podrían ser mayores que los beneficios.

de la avispa que la haya preparado para reconocer y responder a esta nueva razón para actuar.

Las primeras razones para la acción que aparecieron en el universo eran semejantes a las formas platónicas, puros *abstracta* cuya existencia, aunque dependía de la existencia de los luchadores y los buscadores, era independiente del hecho de ser explícitamente reconocida o representada por alguien o algo. Consideremos, por ejemplo, el conjunto de razones, que componen lo que podemos llamar el fundamento lógico de la coloración engañosa de los insectos. El fundamento lógico puede ser muy elaborado: el color y la ubicación precisos de las engañosas manchas (tales como los "falsos ojos" en las alas de muchas polillas) son "dictados" por razones de astucia con el fin de evitar que los predadores localicen la posición de la falena en el momento crítico, entre otros factores. Estos fundamentos lógicos que los pensadores anteriores a Darwin colocaban en la mente del Dios Creador, son considerados ahora como "flotantes", es decir, como un conjunto de razones evaluadas, pensadas y explicitadas por nadie. La sutileza y la complejidad de este pensar sin pensador superan con frecuencia el pensar que piensan los pensadores.[5]

Los animales inferiores como la *sphex* no tienen conciencia, por su constitución, de muchas de las razones que les conciernen. Son similares al personaje secundario de las novelas de espionaje, que desempeña su papel en un proyecto complejo sin tener la menor idea de la verdadera importancia de los hechos en los que se halla involucrado (hasta el desenlace). La CIA y el M15 se manejan con el famoso principio de "saber lo estrictamente necesario" que consiste en explicar lo menos posible y comunicar a los que operan en la propia área sólo lo absolutamente necesario para que puedan ejecutar sus tareas. La Madre Naturaleza es igualmente tacaña a la hora de repartir inteligencia. Cuando se pueden alcanzar "objetivos" más amplios utilizando ejércitos inteligentemente organizados de soldados ignorantes —las hormigas, por ejemplo—, invoca sin compasión la famosa regla de "saber lo estrictamente necesario".

[5] Francis Crick hace referencia a la Segunda Regla de Orgel: "La evolución es más inteligente que usted". Los "fundamentos lógicos flotantes" son definidos en Dennet 1983b. A los intereses y a las razones se los concibe mejor como abstracciones; a menudo son definibles, incluso en los casos en los cuales, basándonos en fundamentos teóricos, creemos que *no pueden* seguirse sistemáticamente. El caso más palmario es la idea de que los intereses de las especies (o de sus grupos) se oponen a los de los individuos (o sus genes). Si el "grupo que selecciona" está equivocado, los intereses de una especie o de ciertos grupos de una especie serán "flotantes" en el sentido más drástico. Los intereses (concedámoslo) son definibles, de modo que podemos determinar cuándo se incrementan (o *les ocurre* un incremento), pero sería erróneo pedirles que desempeñen una función explicativa.

Consideremos, por ejemplo, la estrategia de "saturación del predador" en las tortugas marinas de Ridley. Las tortugas recién nacidas abandonan al unísono la seguridad de sus nidos bajo la arena y participan en una especie de desembarco en Anzio al revés: miles y miles de tortugas salen a la playa y avanzan en medio de una multitud de predadores cuya garra asesina mata aproximadamente nueve de cada diez tortugas antes de que éstas alcancen la relativa seguridad del mar. La estrategia es evidente: ofrecer a los predadores más oportunidades de las que pueden aprovechar. Pero, a diferencia de los bien informados participantes de Anzio, las tortugas desconocen las razones de su sincronizada marcha.

La Madre Naturaleza se atiene al principio de "saber lo estrictamente necesario", pero nosotros defendemos el principio opuesto: nuestro ideal es comprender absolutamente todo, ser conscientes de *todas* las razones que nos incumben, no ignorar lo que es relevante para nosotros y ser guardianes completa y perfectamente informados de nuestros intereses. En esto consistiría la capacidad de elegir las propias acciones según los dictados de la razón.[6]

A menudo, decimos que "la razón dicta" una acción determinada a un actor en una circunstancia determinada. Esto no significa que haya una fuerza personificada, un dictador llamado "Razón" que promulga decretos. Obviamente, se trata de algo abstracto: significa que un problema determinado (considerado en forma abstracta, es decir, independientemente de que alguna criatura lo haya planteado o no explícitamente) tiene una solución (óptima) determinada. El problema queda definido por las circunstancias y por los intereses del actor en cuestión. Pero la imagen subliminal de la vieja y sabia Razón *dictando* lo que hay que hacer ha influido profundamente en el modo en que los filósofos han abordado esta cuestión.[7]

[6] Wolf (1980) discute el caso de alguien cuyo pensamiento está determinado de manera tal que "las razones que *tiene* para ser generoso dependen de las razones que *hay*" (pág. 188). Williams (1981) en "Internal and External Reasons", se refiere a una cuestión similar aunque más específica.

[7] Alentando, por ejemplo, cierta visión de la razón como una "facultad" de la mente que compite con otras; esta visión aparece en el pasaje de Hume:

> Es imposible que la razón pueda tener el efecto ulterior de impedir la volición al dar un impulso contrario a nuestra pasión; y si ese impulso obrara solo, sería capaz de producir la volición. Nada puede oponerse o retardar el impulso de la pasión, salvo un impulso contrario; y si este impulso contrario surge siempre de la razón, esta última facultad debe tener una influencia primaria en la voluntad y debe ser capaz de causar y también de impedir cualquier acto de volición. (*Tratado de la naturaleza humana*, II, III.)

Kant subrayó la diferencia entre hacer lo que la razón dicta y hacer lo que la razón dicta *porque la razón lo dicta*. En el primer caso, uno puede "hacer lo correcto" por mera casualidad o movido por factores externos e irrelevantes. Encontrarse en una circunstancia semejante sería ciertamente un hecho afortunado; sin embargo, es necesario distinguirlo de la buena fortuna que tiene la persona que posee la facultad más amplia y maravillosa de actuar *movida por razones*.

En un estadio intermedio entre el actor que por azar "hace lo correcto" y aquel que lo hace movido por razones justas, se halla el actor que tiende a hacer lo correcto (porque fue diseñado para tender a hacer lo correcto), pero lo hace inconscientemente. El actor intermedio no atiende directamente la sabia voz de la Razón; en realidad, no puede escuchar ni comprender sus dictados. Sin embargo, parece como si el proceso (o el agente) que diseñó al actor respondiera de ese modo a los dictados de la razón. Nuestra *sphex* es un ejemplo. En su entorno natural (cuando no interfieren los biólogos con sus trampas), la *sphex* hace lo que la razón dicta, pues, efectivamente, la razón proclama la máxima de reconocer el terreno ("¡Mire antes de saltar!"). Pero la avispa no lo hace porque la razón se lo dicta *a ella*, sino más bien porque su conducta, en este único aspecto, se aproxima astutamente a la capacidad de dar respuestas propias de lo que podríamos denominar, haciendo violencia a Kant, razón práctica pura.

Veamos otro ejemplo. El zoólogo de Oxford John Krebs investigó un caso notable de lo que podríamos llamar *razón aparente* en los pájaros (Krebs, Kacelnik y Taylor, 1978). Krebs y sus colaboradores colocaron grandes paros (*parus major*), de a uno por vez, en una situación experimental conocida como "el bandido de los dos brazos". Había dos perchas separadas por unos metros que los pájaros podían presionar con el fin de obtener alimento. Los horarios de suministro de alimento se determinaban al azar, pero cada percha tenía un horario distinto. Puesto que se daba por sentado que a los paros les interesaba reunir el alimento de la manera más eficiente posible, su "tarea" consistía en determinar, mediante presiones exploratorias, cuál era la percha que tenía mejor rendimiento medio en relación con sus esfuerzos; por consiguiente, una vez identificada, debían presionar exclusivamente esa percha. Si los paros encontraban el equilibrio óptimo entre la "exploración" y la "explotación", podían maximizar la ingestión de alimento durante su permanencia en el aparato. Por una parte, debían evitar instalarse demasiado pronto en la percha de menor rendimiento y, por la otra, debían evitar prolongar una exploración ineficaz. Las soluciones óptimas de este conjunto de problemas se pueden calcular mediante

un elaborado algoritmo de programación dinámica. En otras palabras, la razón dicta un determinado curso de acción (basándose en hipótesis plausibles sobre los intereses y las circunstancias de los pájaros). Asombrosamente, los pájaros se acercaron mucho a esta solución óptima en el experimento: hicieron realmente lo que la razón dicta. Pero, como subraya Krebs, no hay ningún motivo para suponer que la etiología de sus acciones incluye un cálculo complejo y, menos aun, una comprensión de los principios que dictan la solución óptima (por ejemplo, el principio del máximo de Pontryagin). Frente a un problema similar, una matemática podría encontrar, efectivamente, la solución óptima y, *debido a* su plena comprensión de la misma (a su conocimiento del significado de los propios cálculos), tomarla como fundamento para actuar. Ese sería un ejemplo de un ser movido por la razón. Pero ¿cómo se puede salvar el abismo entre el cerebro de los pájaros y el cerebro de la matemática? ¿Cómo podrían las criaturas más avanzadas (nosotros, por ejemplo) ser capaces de responder directamente a las razones? Para decirlo en términos kantianos, ¿cómo es posible una voluntad verdaderamente racional?

2. Motores semánticos, máquinas de movimiento perpetuo y un generador de intuición defectuoso

Muchos opinan que no hay diferencias de grado entre la "ciega", "mecánica" y sólo en apariencia inteligente conducta de los pájaros o las abejas y la capacidad humana de contemplar y de ser movido por razones, de "capturar" significados y de gozar de una auténtica "intencionalidad". Si el significado no es una propiedad física de las cosas que hay en el mundo, ¿cómo *puede haber* transductores, detectores y reconocedores físicos o mecánicos de significados? El razonamiento subyacente (y rara veces explícito) es simple: los mecanismos físicos de discriminación sólo pueden discriminar propiedades físicas; esto es, pueden clasificar las naranjas según el peso, el color o la forma, pero no según su belleza o según hayan pertenecido alguna vez a un viudo. Los acontecimientos se pueden discriminar físicamente de acuerdo con su duración y su ubicación (o, digamos, por la presencia o ausencia de cierta longitud de onda o por el volumen y el tono de un diapasón). Empero, los acontecimientos no se pueden discriminar físicamente según tengan o no significado. *Nosotros*, sin embargo, podemos discriminar los significados y responder directamente al valor, al sentido o a la relevancia de las cosas que encontramos. En consecuencia, no somos meros mecanismos físicos de discriminación.

Veamos ahora un argumento algo diferente que nos conduce a la misma conclusión. Hacer algo porque la razón lo dicta es ser movido a actuar en esa forma por (la valoración de) una razón. Pero las razones no son condiciones físicas del mundo. Luego, si alguno de nuestros actos es causado por las condiciones físicas del mundo, no es *ipso facto* causado por (la valoración de) una razón. Así pues, la única esperanza de tener una voluntad racional entraña la exención de la causalidad física en nuestra mente.

Se ha insistido tanto en la distinción entre razones y causas que se las ha convertido en antagonistas. Consideremos, por ejemplo, la siguiente observación de Wittgenstein:

> Las causas [*Ursachen*] de nuestra creencia en una proposición carecen en verdad de relevancia en relación con la pregunta de qué es lo que creemos; no así las razones, que están relacionadas gramaticalmente con la proposición y que nos dicen qué proposición es. (Wittgenstein, 1967, i. 437.)

Este párrafo se mantiene en equilibrio inestable; por un lado, se puede interpretar que afirma que los fundamentos o razones de una creencia podrían no ser las causas de que tengamos esa creencia y, por el otro, se puede interpretar que afirma que si una creencia tiene causas no puede ser creída por *razones* —¡lo cual es simplemente falso!—. Dudo que Wittgenstein deseara una lectura semejante; pero, ciertamente, nos invita a hacerla y muchos filósofos han aceptado la invitación con gusto.

Aparentemente, Kant sucumbió a los encantos de este argumento o de alguna de sus versiones. No lograba entender cómo un acto humano podía ser el efecto de causas físicas y, al mismo tiempo, la ejecución de la decisión de una voluntad racional. Como señala uno de sus comentadores, Kant afirma que "tener voluntad significa simplemente poder ser movido por la razón y *no* [la bastardilla es mía] por causas naturales" (Wolf 1973, pág. 216). Así, pues, su visión de la mente se convirtió en un *cul de sac* dualista, ya que concebía a la voluntad como una de las partes de la mente,

El dualismo —la idea de que la mente, a diferencia del cerebro, está compuesta por materia que no está sujeta a las leyes de la naturaleza física— es una visión desesperada que bien merece la desaprobación que se ha ganado. No obstante, si creemos que actualmente podemos explicar los poderes de la mente en términos de facultades puramente físicas del cerebro, debemos admitir que es necesario rebatir este tipo de argumentos. ¿Para qué sirve el cerebro? No para enfriar la sangre, como pensaba Aristóteles; sirve para controlar el cuerpo sobre el cual se ha encaramado, distinguiendo el significado de los obstáculos o los estímulos que dicho cuerpo encuen-

tra. En suma, el cerebro manipula los significados, procesa la información o, como dije antes, es un motor semántico (Dennett 1981a).[8]

Pero, al mismo tiempo, el cerebro es un órgano físico muy complicado que reacciona a los cambios y diferencias físicos en los estímulos. En síntesis, como mecanismo físico, sólo puede ser un *motor semántico*, que responde únicamente a propiedades formales o estructurales.[9] Según la tradicional distinción lingüística, una cosa es la forma o la sintaxis de una proposición y otra el significado o la semántica. Ahora bien, ¿cómo logra el cerebro obtener semántica de la sintaxis? No lo logra.

En sí misma, la sintaxis no determina la semántica. Es una tarea relativamente sencilla diseñar y construir máquinas que verifiquen la acidez o la temperatura inferior a los 40 grados Fahrenheit o la transmisión del S.O.S. Pero, piénsese en diseñar una máquina que verifique la presencia de hostilidad, de amor, de miedo o de escepticismo. A *nosotros* no nos es imposible reconocer, comprender o reaccionar a esas propiedades. Pero es improbable que una combinación cualquiera de propiedades estructurales (y, por ende, fácilmente detectables de manera mecánica), sea equivalente a la hostilidad, al amor, al escepticismo o al miedo. Puesto que el significado no se encuentra, como algún metal raro, en los elementos físicos de los estímulos, ningún proceso alquímico podría destilarlo y reaccionar a él. Luego, el concepto de motor semántico es, en apariencia, similar al de la máquina de movimiento perpetuo: los motores semánticos son, estrictamente hablando, ¡máquinas imposibles!

Pero, entonces, ¿para qué sirve el cerebro si no es un motor semántico? El cerebro sólo se aproxima al comportamiento del motor semántico ideal, puro. El motor semántico perfecto, la perfecta voluntad racional kantiana, está libre de toda fricción, infinitamente alerta a los matices del significado y es totalmente invulnerable a la *sphexidad* y es físicamente imposible. Sin embargo, nuestro cerebro resulta un buen sustituto. Está diseñado con astucia para ser, no infinitamente, sino indefinidamente sensibles a los cambios significativos, no perfectamente, sino prácticamente invulnerable a la *sphexidad*.[10]

[8] Haugeland desarrolla con más amplitud el concepto en "Semantic Engines: an Introduction to Mind Design" (Haugeland, 1981).

[9] Véase también la discusión de Fodor sobre lo que llama la "coacción de la formalidad" (Fodor 1980, reimpreso en Haugeland 1981 y en Fodor, 1981).

[10] Hofstadter (1982b) observa con agudeza las manifestaciones de nuestra inevitable e infinita susceptibilidad ante los absurdos atolladeros en los que quedamos atrapados, lo mismo que la *sphex*. De manera persuasiva, nos dice que nuestros atolladeros, a diferencia de los de la *sphex* se reconocen tanto en nuestros amables desvíos, como en

¿Cómo se las ingenia el cerebro para aproximarse a la eficiencia de un motor semántico perfecto? Ya hemos visto en el caso de la *sphex* cómo una pizca de mera reactividad física inteligentemente diseñada puede causar efectos correctos en ambientes normales, logrando que la *sphex* actúe como si respondiera directamente a un interés, a una razón que le es propia. ¿Se puede lograr una amplificación de tales facultades mediante la combinación y repetición de esos mecanismos básicamente semejantes a los de la *sphex*? La elaboración de dispositivos de reconocimiento de configuraciones cada vez más exquisitos y sensibles puede rastrearse en la filogenia y los adelantes en inteligencia artificial comienzan a revelar los poderes y las limitaciones de esos indirectos e imperfectos detectores de significado. Pero, como señala Hofstadter, el progreso verdaderamente explosivo se da cuando la capacidad de reconocer configuraciones se vuelve sobre sí misma. La criatura que no sólo es sensible a las configuraciones del entorno, ha dado un paso fundamental.[11] A diferencia de la *sphex*, puede darse cuenta de que está atrapada en un brete, en un circuito inútil y, en consecuencia, puede saltar y librarse.

Uno describe las configuraciones en un lenguaje —tanto sintáctico como semántico— por medio de enunciados expresados en un *metalenguaje*. Generalizando un poco más, digamos que uno se "vuelve *meta...*" cuando representa las propias representaciones, reflexiona sobre las propias reflexiones y reacciona ante las propias reacciones. La facultad de reiterar las propias facultades, de aplicar todos los trucos que uno tiene a los propios trucos, es una manera muy conocida de abrirse camino en muchos dominios: un aluvión de procesos que conducen desde la actividad estúpida hasta la actividad compleja. (Hofstadter [1979] ha estudiado exhaustivamente el poder de este proceso de reiteración.)

Así pues, 1) el método ciego de ensayo y error de la selección darwiniana crea 2) organismos cuya ciega conducta de ensayo y error está sujeta a la selección por medio del refuerzo creando 3) con-

---

nuestro estilo artístico y cognitivo; en ocasiones, esto suele ser moderadamente incapacitante, pero, de todos modos, un precio bajo a pagar por nuestra habilidad para pensar rápidamente bajo la presión real del tiempo.

[11] Como Hofstadter (1982b) también señala, el hecho es claramente reconocido por Lucas en el clásico argumento antimecanicista: "Minds, Machines and Gödel" (Lucas 1961). Luego de llegar a esta conclusión, Lucas la desperdicia al cometer el tradicional error filosófico de promover el absolutismo. Sin razón aparente, insiste en que los seres humanos son perfectamente disímiles respecto de la *sphex* y perfectamente coherentes; es decir, perfectos motores semánticos. Como veremos, esta exigencia de perfección absoluta y de invulnerabilidad conduce a menudo, a quienes construyen teorías filosóficas, atribuir propiedades mágicas (esto es, imposibles) a los agentes humanos.

ductas "aprendidas" que generan una multitud de 4) oportunidades de aprendizaje a partir de las cuales 5) las más eficaces pueden ser elegidas "ciega" pero acertadamente, creando 6) una capacidad más precisa para generar 7) nuevos candidatos para una "deliberación" no tan ciega y 8) la eventual selección, elección o decisión de una acción "basada en" esas deliberaciones. Eventualmente, se crea la poderosa "ilusión" de que el sistema responde directamente a los significados. Se vuelve una imitación cada vez más precisa del motor semántico perfecto (la entidad que escucha directamente la voz de la razón), porque fue diseñado para poder mejorarse en este aspecto y para rediseñarse indefinidamente a sí mismo.

La posibilidad de reiteraciones múltiples indica que hay diferentes clases de reflexión, con facultades diferentes. En el extremo más simple del espectro, se halla la reacción simple a algunas de las propias reacciones. Un dispositivo automático para abrir puertas, "que cuenta las veces que funciona" y se desconecta después de un cierto número de ciclos, es tan simple como dicho sistema. Mecanismos similares —meros protectores contra el exceso de repetición— aparecen en animales no humanos, pero producen resultados muy poco sorprendentes. Una vez, yo tenía un perro que adoraba recoger las pelotas de tenis en la pista; cuando se enfrentaba con dos pelotas, no podía sostener a ambas al mismo tiempo, se ponía a dar vueltas en el lugar atrapando una y dejando escapar la otra, hipnotizado por su tendencia a preferir atrapar la pelota en lugar de mantenerla en la boca. Y así continuaba —quizá durante docenas de ciclos— hasta que sonaba un "clic" en su cerebro y se desconectaba ese comportamiento.

Mucho más alto en el espectro se halla la *representación* de algunas de las propias reacciones, sea mediante la expresión de un lenguaje común o en un sistema de representaciones comunes, sea mediante estados internos que a su vez tienen bastante de las propiedades funcionales de los vehículos de representación pública digna de ese nombre.[12] (En la próxima sección examinaremos cómo podría aparecer un sistema con la facultad de autorrepresentación explícita de algunos de sus estados.)

Un sistema capaz de reflexionar así sobre sus actividades no puede reflexionar sobre todas sus actividades al mismo tiempo.

[12] Cómo y por qué trazar una línea entre los estados o acontecimientos que son representaciones y los que no lo son, es un importante tema que he tratado extensamente en otros libros (por ej., Dennett 1978a, 1981b, 1982a, 1983a). En esta ocasión, intentaré soslayar la controversia, apelando sólo a las representaciones cuando su presencia e injerencia en el problema sea obvia para todos, yo incluido. Puesto que soy uno de los escépticos más extremistas respecto de las representaciones mentales, debería ahorrarme el empleo de esta táctica.

Puede reflexionar sobre sus reflexiones, pero tarde o temprano se quedará sin movimientos, ni espacio ni tiempo. ¿Puede existir un observador de sí mismo perfecto? No, si éste no es mágico o infinito. Esto se muestra en teoría de la computabilidad mediante la demostración de que el "problema de la parada" no tiene solución: no existe ningún programa que pueda examinar otro programa y determinar si contiene o no infinitos ciclos (Hofstadter 1982b, págs. 27-28). Ahora bien, ¿hay algo imperfecto que sea bastante bueno para nosotros? Si no somos perfectos, entonces en algo nos parecemos a la *sphex*. Esto significaría que, en principio, nos pueden colocar en una situación ridícula donde manifestemos nuestra marginal e invisible irracionalidad (o, mejor aun, nuestra a-racionalidad o *"sintactitud"*, término sugerido por Hofstadter). Pero esto difícilmente nos provoque conmoción o disgusto. ¿Quién de nosotros necesita los argumentos de la finitud física y de la teoría de la computabilidad para convencerse de que no es un perfecto "extractor de significados"?

En la medida en que concentremos la imaginación en los casos de racionalidad imperfecta y de falta de entendimiento tan dramáticos como el de la *sphex*, es bastante sencillo convencernos de que los seres humanos somos, radical y cualitativamente, una clase diferente de seres. Por otra parte, la razonable inclinación de los filósofos por los ejemplos simples y fáciles de seguir los conduce una y otra vez a concentrarse sólo en esos casos simples cuando contrastan *"el hacer algo por una causa"* con *"hacer algo por una razón"*. En filosofía, toda la literatura acerca de "razones y causas" está infestada por este tipo de comparaciones espurias.

Considérese, por ejemplo, el efecto sutil creado por MacIntyre: "Si la conducta de un hombre es racional, no se lo puede determinar por el estado de sus glándulas o por cualquier otro factor causal precedente" (MacIntyre 1957, pág. 35).[13] ¿Por qué las glándulas? Porque no hay duda de que la clase de cansancio que puede provenir de las glándulas será una suerte de ráfaga, un empellón en una dirección única, un "cambio ciego" en la miríada de configuraciones interconectadas, pero de ningún modo la sutil, delicada y casi indescriptible elección de un estado posterior. He descubierto que existe una palabra de uso común referida a este tipo de causación burda; a menudo se la llama "psicológica" Consideremos la siguiente reflexión extraída de una entrevista a un profesor universitario (no a un filósofo) sobre el tema del libre albedrío:

[13] Este pasaje se discute con más detalle en Dennett 1973, reimpreso en Dennett 1978a. Para otra versión de esta espuria comparación, véase Anscombe 1957, pág. 24 y Dennett 1969, págs. 162-163.

> Es parte de nuestra responsabilidad tener conciencia de lo que hacemos y por qué lo hacemos. Es la única oportunidad de ejercer nuestro libre albedrío y de ser capaces de una elección verdadera en lugar de *dejarse empujar por determinantes psicológicos* [la bastardilla es mía] (Doan y Macnamara, en preparación, manuscrito, pág. 7.13).

Tomar una decisión luego de una exhaustiva consideración de los pros y los contras no es lo mismo que "ser empujados" por determinantes "psicológicos". Este uso de la palabra "psicológico" es similar al uso que los niños norteamericanos le dan a la palabra "mental" (como en la expresión: "¡Ese niño es tan estúpido que debe ser mental!". Evidentemente, hay aquí una elipsis de "retardado mental" o de "enfermo mental"). El uso del término "psicológico" probablemente se debe al sitio de privilegio que ocupa la psicología *clínica* en la imaginación popular. La palabra significaría, más bien, *psicológicamente patológico.* [14]

El hecho de que una acción o una decisión tenga una causa no implica que no se realice también por (buenas) razones. Los filósofos ocasionalmente se valen de este hecho; por ejemplo Nagel, cuando señala que es dable *esperar* que aun cuando las acciones están determinadas por causas, "no están determinadas por causas si ello implica que no están justificadas por buenas razones" (Nagel 1981). No se trata en absoluto de una esperanza lógicamente perdida, pero los filósofos suelen pasarla por alto. Existe incluso un famoso aforismo que afirma rotundamente que la esperanza está perdida: *"tout comprendre c'est tout pardonner"* [comprender todo es perdonarlo todo]. ¿Por qué pensamos que enuncia una verdad?[15]

Porque, quizá, cuando las personas reflexionan sobre el aforismo e imaginan que tienen un conocimiento absoluto de la etiología de algún acto, se encuentran imaginando alguna cosa (de manera esquemática) con un ancestro causal demasiado simple y directo para ser un candidato plausible para un acto responsable. Puesto que los únicos procesos causales que se pueden comprender completamente de una ojeada (el tipo de ojeada que arrojamos a esos temas, en los casos imaginarios) son procesos causales absolutamente simples, intuimos entonces que el conocimiento causal excluye

[14] También el término "biológico" se usa de esa manera. Cuando se alude a algo crudamente biológico. A veces, uno encuentra la expresión "determinismo biológico", donde, según el contexto, debería decir "determinismo genético". Sobre el fantasma del determinismo genético, véase Dawkins 1982, pág. 29.

[15] Por cierto, no todos lo hacen. Strawson me contó que el tema surgió hace años en un grupo de discusión de metafilosofía, en Oxford, y que Austin había respondido: "Eso es absolutamente falso. La comprensión sólo podría agregar desprecio al odio." (¡El incisivo escalpelo de Austin en uno de sus mejores momentos!)

la evaluación racional y la asignación de responsabilidad. En este caso se ha abusado del generador de intuiciones, que es el que ha hecho todo el trabajo.

Cuando se piensa en la causación, se tiende a pensar en los casos de laboratorio, donde es posible obtener un efecto simple, reiterable y simple en condiciones controladas. O bien se piensa en los casos elementales de causación ordinaria: las bolas de billar de Hume, las chispas que provocan explosiones, una cosa grande y visible chocando contra otra cosa grande y visible, etc. Sabemos que un examen más minucioso nos revelaría que cada rincón de nuestro mundo rebosa de inextricables e indescifrables telarañas de causalidad, pero tendemos a ignorar el hecho. Así, cuando alguien es llevado a creer en esto o aquello, imaginamos que ha sido llevado *por la fuerza* a ese estado de creencia, homologándolo con la bola de billar que se hace rodar en determinada dirección o con el agua que se hace hervir en el tubo de ensayo.

Pero, al confiar en esa analogía, tendemos a ignorar una diferencia fundamental: los tubos de ensayo y las bolas de billar son ciegos y sordos. Las personas y los animales superiores están diseñados para ser extremadamente sensibles a prácticamente todo lo que pasa a su alrededor. Es imposible lograr que un ser humano despierto y normal se encuentre en el mismo "estado cognitivo" (además del mismo estado microscópico) en dos ocasiones diferentes, por la simple razón de que el tiempo pasa y, a menos que esté en coma, uno *percibe* y, en consecuencia, cambia constantemente.[16]

No obstante, cuando reflexionamos sobre las implicaciones del determinismo (por ejemplo, la creencia de que nuestras creencias son causadas), nos volvemos casi invariablemente hacia la concepción que ignora (o deliberadamente descarta) el aparato de recolección de información que es nuestra más eficaz interfase causal con nuestro entorno.[17]

Consideremos, por ejemplo, las fuerzas que han arrastrado la imaginación de Ayer cuando se refiere a este generador de intuiciones de ciencia-ficción:

[16] Se verá que este hecho desempeña un papel importante en el análisis correcto de la expresión "podría haber actuado de otra manera" en el capítulo 6. Lucas (1970, pág. 79) repara en la importancia del hecho, pero, como de costumbre, no puede dejar de inflar su posición y llega hasta el punto de hacer un llamado a la "infinitud" humana.

[17] Fue Norbert Wiener en *Cybernetics: On Control and Communication in the Animal Mind and the Machine* (Wiener 1948) quien imaginó la visión contemporánea de los seres humanos como "autómatas acoplados de manera eficaz al mundo exterior, no solamente por el flujo de energía o por el metabolismo, sino también por el flujo de impresiones y mensajes que reciben y por las acciones de los mensajes que se emiten" (pág. 54).

> Supongamos que se desarrollaran las teorías fisiológicas o psicofisiológicas pertinentes y que pudieran utilizarse para realizar importantes y acertadas predicciones en la vida cotidiana. Si esto les parece muy fantasioso, supongamos entonces que la teoría del condicionamiento ha evolucionado tanto que es posible ahora implantar deseos, creencias y rasgos de carácter en los seres humanos, y también deducir cómo se comportará una persona sometida a ese proceso en una determinada situación. Supongamos que vivimos en un régimen donde esos poderes se han ejercido sobre nosotros, digamos, desde la más tierna infancia. (Ayer 1980, pág. 9)

Una visión escalofriante, sin duda, que Ayer ha usado para "extraer" la siguiente intuición: "nuestras actitudes en las relaciones interpersonales", la percepción de nosotros mismos y de los demás como agentes responsables, culpables y dignos de amor, no soportarían el conocimiento que tendríamos entonces del comportamiento de los agentes. Me atrevo a decir que no lo soportarían *en las circunstancias imaginadas.* En esas circunstancias (para no ser "demasiado fantasioso") la causalidad real de las creencias y deseos imaginados se ha vuelto radicalmente más cruda y los agentes han sido forzados a padecer supuestas "creencias, deseos y rasgos de carácter" mediante un proceso de condicionamiento que se presume lo bastante poderoso para anular el delicado proceso de recolectar información y evaluarla, que es nuestra manera normal de tener creencias y deseos.

Vivir en un mundo semejante "sería lo mismo que convertirse en el espectador de una obra teatral en la que también se participa, sin otra opción que la de desempeñar el papel asignado" (Ayer 1980, pág. 10). Pero ¡aguarden un momento! ¿Si comenzamos a murmurar y a reírnos por lo bajo desbaratando la obra y demostrándonos mutuamente la futilidad de la trama? ¡Seríamos injustos con el ejemplo de Ayer! No se supone que no haremos una cosa así. ¿Por qué no? ¿Estipuló Ayer que en ese mundo imaginario las personas ya no perciben las cosas ni reaccionan de acuerdo con ellas? Los científicos malignos que mueven los hilos de ese mundo nos pueden espiar a través de un vidrio blindado, sin inmutarse por nuestra rebelión. Ayer nos ha convertido en criaturas mucho más simples. (Véase el método operativo en el generador de intuiciones de Ayer con el método usado por Robbins en el cuento de la aldea tirolesa, citado en el capítulo 1.)

3. *Reflexión, lenguaje y conciencia*

En cierta ocasión le preguntaron a Hobbes: "Si la voluntad de las personas está determinada, ¿para qué presentarles razones?";

Hobbes respondió: "Porque de esa manera pensamos que los induciremos a tener la voluntad que no tienen" (Hobbes, *Works*, vol. V, pág. 52). Las personas normales tienen la suerte de poseer una voluntad determinada (si lo está) por un complicado proceso causal de percepción y de raciocinio que a menudo incluye —como lo señala Hobbes— descargas causales de comunicación. A las personas se les presentan razones y a menudo esas razones las mueven a actuar, afortunadamente para ellas. Las personas, pues, se mueven por razones y lo hacen de una manera esencialmente distinta de la causalidad "ordinaria". Pero me pregunto si esa "manera distinta" no lo será en un sentido que he pasado por alto o acaso escamoteado. Hemos esbozado ya un sinnúmero de procesos de autoexamen afirmando que esos procesos nos apartan de la *sphex* y de sus pares y nos acercan a otras criaturas que son, como nosotros, indefinidamente sensibles al valor de las cosas que suceden en el entorno. Esas criaturas son aproximaciones fidedignas a los "auténticos" motores semánticos. Sin embargo, a los ojos de los filósofos absolutistas, tales entidades serían meros simulacros y de ninguna manera algo real.[18] Además del dogma tradicional, ¿qué razones hay para reiterar una y otra vez que somos *nosotros* lo realmente verdadero? Si somos tan buenas aproximaciones, fallarían entonces todas las pruebas empíricas del papel tornasolado.[19]

¿Qué ocurriría si realizáramos una prueba introspectiva de este tipo? Uno se siente tentado a afirmar que nuestra conciencia de las razones nos separa del resto de las criaturas. Al esbozar la secuencia que franquee el abismo entre nosotros y la *sphex*, ¿no habré franqueado subrepticiamente otro abismo: el que separa lo consciente de lo inconsciente? Cuando hablamos del reconocimiento de configuraciones en el caso de la *sphex*, o de las decisiones en el caso de los paros (*parus major*), necesitamos, aparentemente, encerrar entre comillas de espanto a nuestros términos mentalistas. Es tentador insistir en que son meras analogías mecánicas del reconocimiento, la decisión y la facultad de comprensión genuinos. Las facultades de "comprensión" más sensibles hasta aquí descritas constituyen la especie de lo que podríamos llamar comprensión *meramente conductal*. Se supone que este tipo de comprensión se

[18] Véanse Searle 1980 y Nagel 1974. Sobre el absolutismo de ambos autores, véase la discusión de Richard Rorty (1982) acerca de lo que llama su esencialismo, y mi contestación (Dennett 1982c).

[19] Lucas pensó que había encontrado la prueba del papel tornasolado, aplicando el teorema de incompletud de Gödel combinado con la idea de una máquina de Turing (Lucas 1967 y 1970). Véanse mis refutaciones (Dennett 1972 y 1978a, cap. 13) y Hofstadter (1979).

opone a la comprensión auténtica y consciente que hallamos en los seres humanos.

No está claro, sin embargo, qué agregaría esta propiedad (familiar pero no analizada) de la conciencia a las facultades de los sistemas de comprensión "meramente conductal". Quizá la enorme diferencia cualitativa sea otra ilusión provocada por el hecho de conceder demasiada atención a los extremos y muy poca a las complicaciones intermedias. Es verdad que los teóricos que se ocupan de estos problemas se han asentado firmemente en una de las dos concepciones. Los filósofos han adoptado tradicionalmente un abordaje puramente introspectivo, que presupone la conciencia; por el contrario, los biólogos, los científicos de la computación (y los científicos que modelan estos procesos) tienden a un abordaje puro, de "ingeniería", adoptando la perspectiva de la tercera persona; esto es, sin involucrarse en los mecanismos que describen. Estos científicos subrayan expresamente la ausencia de conciencia (o de algo sospechosamente parecido a la conciencia) en sus argumentos.

De igual modo, las "ideas de la reflexión" desempeñaron un papel fundamental en la teoría de la mente de Locke; todas las "ideas de la sensación y la reflexión" estaban al parecer presentes en el conocimiento consciente del agente humano. Locke tenía un agudo sentido del valor de la actividad en los metaniveles, y en este hermoso pasaje fue tan audaz como para definir el libre albedrío según esos términos:

> Pues teniendo la mente en la mayoría de los casos, como se ve en la experiencia, el poder de *suspender* la ejecución y satisfacción de alguno de sus deseos, y así de todos, uno tras otro, es libre de considerar los objetos de éstos, examinarlos por todos lados, compararlos con otros. En esto reside la libertad que tiene el hombre; y por no usar su derecho viene toda la variedad de errores, equivocaciones y faltas en los que incurrimos en la conducción de nuestra vida y en nuestros esfuerzos por procurarnos la felicidad; y así precipitamos la determinación de nuestra voluntad y nos comprometemos demasiado pronto, antes del debido examen. Para evitar esto, tenemos el poder de suspender la prosecución de ese o aquel deseo; como cada uno lo experimenta cotidianamente dentro de sí mismo. Esto me parece a mí la fuente de toda libertad; en esto parecería residir lo que se llama (para mí de manera impropia) *libre albedrío*. Durante la suspensión del deseo, tenemos oportunidad de examinar, considerar y juzgar lo bueno y lo malo de lo que haremos; y cuando, basándonos en el debido examen, juzgamos que hemos cumplido con nuestro deber y hecho todo lo que podíamos o debíamos hacer en prosecución de la felicidad; y no es una falta sino una perfección de nuestra naturaleza desear, poder y actuar de acuerdo con el último resultado de un análisis justo. (*Ensayo sobre el conocimiento humano* II, XXI, pág 48.)

Como señala Locke, la mayor parte de nuestro pensar no está referido a los objetos de nuestros deseos sino a nuestros deseos de los objetos. Esto nos lleva a preguntarnos si hay objetos dignos del deseo, "consideradas todas las cosas". (Como veremos en el capítulo 4, es importante que no podamos "considerar todas las cosas"; ni siquiera "todas las cosas" relevantes.) Tenemos, también, la facultad de pensar acerca de nuestras creencias y de preguntarnos si son lo que deberían ser.

Este *pensar acerca de las creencias y los deseos*, este reflexivo metapensar, ¿es consciente o se trataría de un pensar meramente conductal, semejante al tipo de "procesamiento de información reflexivo" que puede ejecutar un ordenador? Muchas cosas se derivarían de la respuesta, pues algunas de nuestras intuiciones más poderosas sobre la libertad dependen de la conciencia. Cuando tomamos conciencia de las razones, las reconocemos (a diferencia de la *sphex* que supuestamente se comporta como si reconociera sus razones). De ese modo, la capacidad de reconocer conscientemente los motivos es una condición al parecer necesaria para la verdadera libertad. Strawson describe el caso de un psicoanalista que trató con éxito a un paciente, y señala que "naturalmente podemos decir" que el psicoanalista "ha restaurado la libertad del agente". Pero aquí "restaurar la libertad" significa que la conducta del agente será inteligible en términos de propósitos inconscientes (Strawson 1968, págs. 90-91). Aun si podemos encontrarle un sentido a cierto tipo de pensamiento inconsciente, ¿tiene sentido la idea de un pensamiento inconsciente acerca de deseos y creencias? ¿Podemos ser autoconscientes, pero sólo inconscientemente autoconscientes? ¿O bien el abismo entre los procesos inconscientes y los conscientes se franquea cuando se llega a los sistemas capaces de tratar como objetos de "inspección" algunos de sus estados de "creencia" o "deseo", sometiéndolos a los mecanismos de reconocimiento de configuraciones? La pregunta moviliza sentimientos e intuiciones. Algunas personas dicen saber, en el fondo del corazón, que ninguna cantidad de cuasicomprensión meramente conductal podría dar por resultado la comprensión consciente o la auténtica intencionalidad.

Otros, incluyéndome a mí mismo, estamos igualmente seguros de que la "auténtica conciencia" o la "verdadera intencionalidad", cuando se manifiesta, debe ubicarse en la cúspide de la pirámide de los procesos físicos naturales. Trataré de hacer más atractiva la idea, desarrollando algunos de mis generadores de intuición, aunque eso signifique apartarme del tema del libre albedrío rumbo a los campos de batalla de la filosofía de la mente; pero a menos que haga algo para reconocer y aliviar el escepticismo en este aspecto, sospecharán que me extralimité.

La creación y el desarrollo de la inteligencia es uno de los resultados más formidables de la evolución, pero no todas las cosas vivientes se comprometen en la batalla de talentos o, en todo caso, no con el mismo grado de compromiso. En la evolución, los "puntos de elección" se hallan, por así decirlo, entre la *Línea Maginot* (que consiste en "cavar la trinchera" y optar por la inmóvil y relativamente acorazada invulnerabilidad de las plantas, las almejas y otros seres vivientes cuya conducta revela apenas un mínimo de destreza) y la *guerra de guerrillas* (la táctica de ocultamiento y persecución contra los otros jugadores del entorno). Las especies que "optan" por la guerra de guerrillas participarán en una carrera armamentista cognitiva contra las otras criaturas que también *eligieron* la movilidad y la inteligencia en lugar de la coraza. Nótese que la capacidad meramente conductal (o de "procesamiento de la información") de "preguntarse a uno mismo" por la genealogía de las "creencias" y por la coherencia de los propios "deseos" (la capacidad que Locke encomia, sin la presunción de conciencia) significa un avance importante en la carrera armamentista cognitiva. Incluso la imperfecta capacidad de "evaluar" algunos de los propios estados cognitivos y conativos implica una gran diferencia. Sin ella seríamos ingenuos y, por ende, vulnerables a la manipulación de cualquier agente (secreto) que descubriese la manera de inducirnos a tal o cual creencia.

La tradición ha valorado tácitamente el poder del examen reflexivo, como lo demuestra la literatura sobre el libre albedrío referida a experimentos mentales protagonizados por manipuladores inescrupulosos. En tales experimentos, es fundamental que la víctima ignore la existencia del manipulador. Si el autoexaminador se enterase de las anormalidades provocadas en la causación de sus estados cognitivos (y, especialmente, en el origen e intención de esas anormalidades), destruiría de inmediato los planes del impostor. El proceso de deshacer un engaño se lleva a cabo de dos maneras: la ilusión subsiste, pero uno aprende a descalificarla y a desecharla (las llamadas ilusiones "estables"), o bien la pérdida de la inocencia lo vuelve a uno incapaz de experimentar la ilusión (o, mejor aun, uno se vuelve más o menos invulnerable a la ilusión).

La tarea del maligno embaucador es bastante difícil, aun cuando las víctimas sean ingenuas. Los órganos de nuestros sentidos más avanzados están admirablemente diseñados para ser lazos causales de alta fidelidad entre el mundo y las creencias sobre el mundo que se originan en la percepción, de modo que es prácticamente imposible engañar a una víctima a la que se le permiten mínimas exploraciones: mover la cabeza, los ojos o cambiar su posición de ventaja (Gibbon 1966 y 1976). Si la víctima, además,

tiene el hábito de reflexionar periódicamente sobre la genealogía y la coherencia general de sus creencias y deseos, la tarea del embaucador maligno se vuelve virtualmente imposible.

¿Cómo se puede desarrollar la facultad de autoexamen y de autocrítica? Deseamos convencernos de que tal desarrollo es posible. Podemos posponer los detalles verdaderos hasta que la ciencia se encargue de descubrirlos; por ello, todo lo que necesitamos en este momento es "Sólo un cuento" plausible.[20]

Había una vez unas criaturas que eran totalmente inconscientes, aunque poseían órganos de los sentidos que funcionaban en plenitud y los "informaban" sobre las condiciones del mundo. Su vida, sin embargo, era muy compleja; tan compleja que cuando uno les aplicaba el principio de "saber estrictamente lo necesario", se daba cuenta de que necesitaban "saber" (en el sentido "meramente conductal") mucho sobre sus actividades. En particular, tenían problemas de coordinación con los otros miembros de la especie y la solución aparentemente óptima de esos problemas requería formas rudimentarias de "comunicación". Esta comunicación era, en cierto modo, semejante a la danza de las abejas y a otras formas de transmisión de mensajes entre los insectos; pero se diferenciaba de ellas porque, cuando una criatura se comunicaba con otra, "sabía" de qué estaba hablando y por qué; es decir, no se comunicaba meramente a la manera de la *sphex*, cada vez que aparecía un "disparador" en el entorno (como la abeja "llevada" a danzar por la presencia de néctar en las inmediaciones).

Estas criaturas (o sus antepasados) "advirtieron" que se podían obtener mejores resultados cuando "discriminaban" diferentes oyentes en diferentes ocasiones según lo que ambas partes "supieran", "creyeran" y "desearan" (a su manera meramente conductal).

La criatura Alf, por ejemplo, no se molestaría en tratar de que la criatura Bob "creyera" que no había más alimento en la cueva, si Alf "creyera" que Bob ya "sabía" que no había más alimento en la cueva. Y, si Bob "pensara" que Alf "quería engañarlo", Bob podría "descreer" de lo que Alf decía.

Existen en la actualidad programas de inteligencia artificial (IA) que modelan en profundidad la estructura organizacional de los sistemas que deben planear interacciones "comunicativas" con otros sistemas, basándose en el "conocimiento" acerca de lo que ellos mismos "saben" y "no saben", de lo que "sabe" y "no sabe" el sistema interlocutor, etc. Podemos suponer que en el experimento mental,

[20] Sobre el abuso de "Sólo un cuento", véanse Humphrey (1982) y el comentario y la respuesta en Dennett 1983b.

nuestros imaginarios progenitores no eran más conscientes de lo que serían los robots de IA. (No tengo la menor idea de qué puede significar esto para la mente de aquellos que hacen hincapié en esto; en este momento, mi estrategia consiste en concederles cualquier punto.)

A veces sucedía que cuando una criatura se empantanaba en algún proyecto, solía "pedir ayuda"; en particular,"pedía información". La mayor parte de las veces el oyente le respondía "comunicando" algo cuyos efectos sacaban a la criatura del atolladero o la inducían a "ver" la solución del problema. Para que esta práctica se ganase un lugar en la comunidad, era necesario que aquellos que preguntaban pudiesen actuar recíprocamente, desempeñando en ocasiones el papel de oyentes y respondiendo a la súplica. Esto es, debían tener la capacidad (meramente conductal) de proferir expresiones "útiles" ante las expresiones de "súplica" proferidas por los otros. Cuando se le preguntaba a un sistema algo que "sabía", la pregunta podía provocar el efecto benéfico y normal (pero con excepciones) de inducirlo "a decir lo que sabía".[21]

Hasta que una vez la criatura percibió que se había producido un "inesperado" cortocircuito en este nueva institución social: "pidió" ayuda en una circunstancia inadecuada, donde no había oyentes que pudieran escuchar y responder a su requerimiento, ¡salvo ella misma! Cuando la criatura oyó su propia súplica, la estimulación provocó la clase de respuesta "útil" que hubiese provocado la súplica de otro, ¡y, para su delicia, la criatura comprobó que se había inducido a responder su propia pregunta!

¿La actividad de preguntarse a sí mismo es menos inútil que la actividad de dejarse a sí mismo una propina cada vez que uno se sirve un trago? Mientras adhiramos al "perfeccionismo ingenuo" (Dawkins 1980) acerca de la mente o del yo y lo consideremos, como Descartes, como un todo que se comunica consigo mismo e indivisible, es difícil imaginar que tales actividades reflexivas puedan servir a algún propósito.[22] Piénsese en todas las ocasiones en las que uno se recuerda, se encomienda, se promete, se reprende y se aconseja a sí mismo. Seguramente esta autoadministración debe surtir algún efecto, si se mantiene con tanta firmeza en nuestro repertorio.

[21] Los problemas técnicos relacionados con la evolución de la comunicación no son triviales pero tampoco insolubles. Se puede suponer que un espíritu cooperativo, de mutua ayuda, puede tener cierto valor para la supervivencia, o bien transformarse en un sistema estable una vez que ha surgido. Véase Dawkins 1982, págs. 55 y siguientes.

[22] Los epistemólogos y filósofos de la mente han pasado por alto y subestimado la utilidad de tales prácticas; un hecho sorprendente puesto que, como Powers (1978) señala, ése es el tema central aunque encubierto del *Menón*, de Platón.

¿En qué condiciones sería útil la actividad de preguntarse a sí mismo? Todo lo que necesitamos suponer es que existe una división en compartimientos y una comunicación *interna*, imperfecta entre los componentes del sistema cognitivo de una criatura, de manera que un componente puede necesitar el *output* de otro, pero puede ser incapaz de dirigirse directamente a él. Supongamos que la única forma de lograr que el componente A realice su tarea es induciéndolo a la acción mediante cierta clase de estímulo que normalmente proviene del exterior, de otra criatura. Si un día uno descubre que puede desempeñar el papel de esa otra criatura y obtener buenos resultados mediante la autoestimulación, la práctica marcará una nueva señal de comunicación entre los propios componentes internos, una señal que se desplazará por el espacio común de las ondas de aire y de la acústica.[23] Para decirlo sin ambages, empujar la información a través de los propios oídos y del aparato auditivo puede estimular las conexiones que está buscando, desatar los mecanismos de asociación correctos y empujar hasta la punta de la lengua el auténtico bocado mental. Entonces, uno lo dice, se lo oye decir y encuentra la respuesta que esperaba.

Numerosos estudios y experimentos sobre la afasia y otros desórdenes mentales muestran con bastante evidencia que la producción del lenguaje no es la imagen especular de la comprensión del lenguaje y viceversa. Cuando escuchamos y comprendemos un enunciado no lo absorbemos a través de la misma maquinaria que se usa para formular y proferir un enunciado, esto es, no se trata de la misma maquinaria funcionando en sentido opuesto. En consecuencia, no hay razones para suponer que el proceso de formular, proferir, enunciar, escuchar y comprender una oración nos dejaría en el mismo sitio de donde partimos cognitivamente. Las tareas cognitivas "automáticamente" subcontratadas en el transcurso de la generación y la comprensión de una oración parecerían ser, justamente, apropiadas para revolver los bolsillos dormidos del conocimiento, que acaso contienen la pieza faltante del rompecabezas del momento.

Pues bien, en "Sólo un cuento", las criaturas se habituaron a hablarse a sí mismas en voz alta y descubrieron que, a menudo, obtenían buenos resultados (en realidad, lo suficientemente a me-

[23] Recientes experimentos con personas con el "cerebro dividido" (esto es, cuyo cuerpo calloso ha sido seccionado, rompiendo la vía normal de comunicación entre los hemisferios) revelan en forma dramática la habilidad del cerebro para buscar y explotar nuevos canales de comunicación. (Nótese que es al cerebro o hemisferio cerebral —no a toda la persona— a lo que se atribuye el inteligente descubrimiento. La persona es completamente inconsciente de los trucos de los que se vale el cerebro para restablecer la comunicación. Véase Gazzaniga y Ledoux 1978.

nudo como para reforzar la práctica). Luego, hallaron un atajo eficaz: hablarse *sotto voce*, lo cual los condujo finalmente a hablarse a sí mismos en completo silencio. Este último proceso mantuvo el circuito de autoestimulación pero se desembarazó de los elementos de la audición y de la vocalización periférica, que no eran de gran utilidad. Esta innovación obtuvo el beneficio adicional de lograr una cierta privacidad para la práctica de la autoestimulación cognitiva. Y la privacidad resultaba especialmente útil cuando los miembros capaces de comprender de la especie se hallaban al alcance del oído (pues no debemos suponer que el intercambio de "socorros", que fue la semilla del proceso, era enteramente altruista y para nada competitivo).

Así evolucionaron innumerables conductas silenciosas y privadas en un marco social dentro del cual se producía una comunicación útil. No era necesariamente el mejor de los procesos cognitivos imaginables; si lo comparamos con otros procesos cognitivos inconscientes, fue lento y laborioso pues hubo que echar mano de extensos tractos de maquinaria "destinada" a otros fines (para la producción del discurso audible y de la comprensión). Era tan lineal —limitado a un tema por vez— como la comunicación social a partir de la cual evolucionó y dependía, al menos en el comienzo, de las palabras comunes que componían la práctica social.

Supongamos que un fenómeno semejante evolucionara en otro planeta. ¿Lo llamaríamos conciencia? ¿Incluiríamos a esas criaturas dotadas de una actividad "meramente conductal" entre nuestros conscientes y autoconscientes cofrades? ¿No sería su actividad de procesamiento de información interna un pseudopensar inconsciente, un pensar "meramente conductal"? En el comportamiento exterior apenas se distinguirían de nosotros: cooperadores pero también huidizos; comunicactivos pero también secretos y aptos, en ocasiones, para sentarse a murmurar mirando el aire hasta que algo cognitivo se abra paso.

¿Nos tienta la idea de zambullirnos dentro de nosotros mismos para determinar si la actividad cognitiva es consciente? En caso de que mi generador de intuiciones haya funcionado, usted debería sentirse tentado a concluir que estas criaturas son conscientes.[24] Es todo lo que le pido: ¡alguna tentación!

Pero si aún se mantiene escéptico, observe que no nos limitamos solamente a la actividad de hablarnos en silencio. ¿Por qué las personas pintan y dibujan diagramas para sí mismas? ¿Por

[24] En Hofstadter y Dennett (1981), se pueden encontrar más generadores de intuición que intentar aumentar aún más este efecto.

qué se molestan los compositores en tararear o tocar música para su exclusivo beneficio? (Goodman 1982). Podemos suponer que la criatura en "Sólo un cuento" sería capaz de participar con provecho en actividades internas tales como canturrear o diagramar; e incluso podría extraer beneficios, al igual que nosotros, de un "partido de tenis interior" (Gallwey 1979).

Hay innumerables técnicas de autoestimulación. Así como notamos que acariciarse a sí mismo de cierta manera produce determinados efectos sólo parcial e indirectamente controlables pero definitivamente satisfactorios (y entonces dedicamos tiempo e ingenio a explorar y desarrollar técnicas que produzcan esos efectos deseables), del mismo modo llegamos a reconocer que hablar, pintar, canturrear para uno mismo son prácticas que con frecuencia producen efectos deseables. Algunos son más aptos que otros para estas actividades. La autoestimulación cognitiva es una técnica adquirida y profundamente personal, con muchos estilos diferentes.[25]

Pero si, aun así, intuye que estas criaturas imaginarias no son conscientes, no como usted, ¡así sea!: trataré de cambiar el rumbo. Recuérdese que esta digresión sobre la conciencia se inspiró en el ejemplo de Strawson acerca del psicoanalista que "restauró la libertad de un paciente" haciendo que la conducta de ese agente fuera inteligible "en términos de propósitos conscientes". La apelación a los propósitos conscientes —opuestos a los inconscientes, nocivos para la libertad— indica que, antes del tratamiento, el paciente no tenía libertad porque su famoso "yo consciente" carecía de información acerca de las fuentes de su conducta. Sólo fue libre cuando pudo informarse a sí mismo y llevarlas a la arena de la discusión y la reflexión racional. Pero una vez que nuestras criaturas imaginarias logran que las facultades lingüísticas y la capacidad de manejar información reflexiva funcionen, ¿serían igualmente sensibles a la persuasión e igualmente capaces de comprometerse en una autoevaluación racional? Esto es, ¿estarían equipadas para reaccionar adecuadamente cuando nosotros, como dice Hobbes, "les

[25] El material del párrafo anterior fue extraído de Dennett 1982c. "Sólo un cuento" se inspiró en diversas fuentes, especialmente en la colección póstuma de Ryle, *On Thinking* (Ryle 1979), en Straight (1977) y en Bennett (1976) que contiene un detallado "Sólo un cuento" sobre el primer lenguaje, la charla común y las tribus —especialmente las que hablan este lenguaje, que son los Estólidos y los Condescendientes (capítulos 5-7)—. Si las especulaciones de Bennett se ocupan de la aparición de los rudimentos del lenguaje, estos rudimentos se ocupan a su vez de la aparición de la conciencia humana, según el ingenioso (y menos riguroso) "Sólo un cuento", de Jaynes (1976). Véase también Humphrey 1982.

presentamos razones"? ¿O no depende de esto la libertad de si aflora o no a la conciencia?[26]

4. Comunidad, comunicación y trascendencia

> *Mejor que nuestras ideas y valores,*
> *nuestros síes y noes, nuestros peros y cuandos*
> *broten de nosotros con la necesidad*
> *con que el árbol da fruto.*
> —*Nietzsche,* Genealogía de la moral

La diferencia entre nosotros y la *sphex* es bastante evidente. En realidad, demasiado evidente. Fomenta la clase de exageración absolutista que lleva a los filósofos a defender teorías extravagantes sobre nuestra condición privilegiada y metafísicamente única de estar exentos de la causalidad. La confirmación de nuestro parentesco esencial con la *sphex* tiende entonces a crear una ansiedad difusa respecto de nuestros poderes acaso ilusorios como "agentes racionales". Mi reseña esquemática de las diferencias con la *sphex* y de los procesos naturales que las originaron nos lleva a reiterar el pensamiento de Darwin: en el principio sólo había materia prima relativamente no estructurada ni elaborada; se sucedieron diversas mutaciones y de este proceso surgió algo nuevo, trayendo consigo no sólo nuevas clases de fenómenos concretos sino nuevos *abstracta*. Del mundo inorgánico emergieron, pues, los replicantes y sus razones, originadas en los esfuerzos por replicarse.

A partir de estas ignaras criaturas movidas por tropismos, que sólo imperfectamente "reconocían" y "cuidaban" sus propios intereses, evolucionaron otras capaces de condicionamiento y de aprendizaje, que podían rediseñarse a sí mismas de modo de volverse cada vez más sensibles a sus razones, y que se perfeccionaron, en especial, por la aparición de una facultad que les permitía reaccionar a las estructuras de sus propias reacciones ante las estructuras de la naturaleza. De las rudimentarias prácticas protolingüísticas de algunas de esas criaturas surgió, por mutación oportunista, una actividad reflexiva aún más poderosa: la representación de algunas de estas actividades y el uso de estas representaciones para el autoexamen y la autoevaluación. Pero hay todavía un peldaño más

[26] La tesis fundamental de sus trabajos durante los últimos quince años ha sido, quizá, la afirmación de que tales criaturas *son* conscientes —nosotros somos esas criaturas— y, por cierto, no pienso abandonarla aquí. Sin embargo, mi argumentación sobre la libertad no depende de ella. Por consiguiente, mis lectores incrédulos no necesitan aceptar todos mis generadores de intuición para proseguir con este análisis de la libertad.

alto en la escalera que va desde la avispa hasta la voluntad racional que Kant glorificó.

La emergencia de una población de comunicadores capaces de lenguaje nos suministra materia prima para otro "Sólo un cuento": el origen de la moral. Son muchas las versiones que la filosofía aprueba e incluso celebra pero, hasta donde sé, fue Hobbes el primero en vislumbrar la necesidad y la belleza teórica del cuento. Su versión tiene las debilidades, las virtudes y el extremismo propios de los esfuerzos de un pionero. A comienzos del capítulo 1 expliqué de qué manera la aparición del interés fue, en parte, una cuestión de intereses, *constituidos* por la clase de acciones que las criaturas debieron realizar en el transcurso del denodado proceso de autoconservación. Un proceso similar, esencialmente constitutivo, fue descrito por Hobbes.

Hubo un tiempo, dijo, en que no existía el bien y el mal; la gente vivía en "estado de naturaleza", forzada a participar en la guerra de todos contra todos, y la vida era "solitaria, miserable, sórdida, bestial y corta" (*Leviatán*, 1651). Entonces, algunos de ellos se agruparon e hicieron un contrato y, de ese modo, comenzó la sociedad y, con la sociedad, el *bien y el mal*. No solamente la idea o el concepto de bien y de mal, sino el bien y el mal mismos. Antes del contrato, nadie era moralmente bueno o moralmente malo. Esto no significa que si la gente hacía cosas buenas o malas no supiese lo mala o buena que era. Nada era bueno o malo en un sentido... moral. En aquel entonces, la gente vivía en estado de naturaleza, como viven ahora los leones y los tigres, y por cierto sabemos intuitivamente que en la relación entre animales y salvajes no existen hechos morales.[27] El contrato lo cambió todo y dió origen a un modo de vida diferente con nuevas razones para actuar, así como la aparición de los replicantes dio a luz la primera forma de vida y, con ella, las primeras razones para vivir.

El "Sólo un cuento" de Hobbes es históricamente falso y simplifica excesivamente las cosas desde el punto de vista conceptual. Proclama, muy engañosamente, que la tan inflada conciencia y el lenguaje humanos pudieron haber florecido en un estado de naturaleza hobbesiano, presocial o asocial. La aparición del orden social, del lenguaje, de la conciencia y de la moral debió ser una maraña de evoluciones interpenetrables e interactuantes para la

[27] Sin embargo, nuestras intuiciones no son unívocas. Si lo fueran no necesitarían el tipo de esfuerzos regulares que reciben, por ejemplo, en casi todos los documentales sobre la vida salvaje, donde el narrador nos recuerda, mientras somos testigos de algún hecho terrorífico de la naturaleza (teñido de rojo y con dientes y garras) que éste es el estilo de la naturaleza: ¡también el león tiene que comer!

cual la explicación de Hobbes, considerada como un "modelo", sería en extremo simplista. Pero el cuento de Hobbes consigue aislar un importante anclaje para el pensamiento: las prácticas sociales de defensa verbal, crítica mutua y deliberación no sólo originaron una nueva relación entre la criatura y sus razones (la relación de *considerar* esas razones) sino la creación de nuevas razones a considerar. Como dice Hobbes, presentamos razones a los demás "porque de ese modo pensamos que podremos inducirlos a tener la voluntad que no tienen". La costumbre de presentar razones solamente es efectiva cuando los otros son personas (no pierda el tiempo presentando razones a los árboles, almejas o a los mentalmente incapacitados). De hecho, es tan efectiva que nos permite, y sólo a nosotros entre las criaturas de la Tierra, "trascender nuestra biología"; pero no trascenderla de un modo misterioso, vitalista, opuesto a la causalidad, sino permitiendo que nuestros sistemas de control cognitivo alcancen objetivos que no *pueden ser* genéticamente controlados.[28] Dawkins, en el último párrafo de *The Selfish Gene* (que muchos de sus críticos, al parecer, nunca leyeron) señala muy bien que:

> Tenemos el poder de desafiar a los genes que poseemos de nacimiento ... Podemos incluso imaginar formas de cultivar y alimentar deliberadamente un altruismo puro y desinteresado, algo que no ocurre en la naturaleza [esto es, en el estado de naturaleza hobbesiano] y no ha ocurrido antes en la historia del mundo... Nosotros, los únicos en la tierra, podemos rebelarnos contra la tiranía de los replicantes egoístas. (Dawkins 1976, p. 215.)

Ahí, aunque lleguemos a este planeta equipados con un conjunto de predisposiciones y aunque nos inclinemos de manera innata por un cierto estado de cosas, podemos sin embargo construir vidas que trasciendan esas inclinaciones. Podemos domeñar, anular e incluso, si fuera necesario, reprimirlas en favor de inclinaciones "más altas", que no son menos reales por el hecho de no ser directamente biológicas (es decir, genéticamente heredadas). (El espectro de la sociobiología se cierne sobre quienes no ven que esto es obvio y piensan que la única manera en que la sociobiología podría contribuir a la comprensión del nacimiento de la moral, sería "reduciendo" las normas morales a imperativos genéticos oscuramente imaginados. Temen las explicaciones "biológicas" por la misma razón que temen las explicaciones "psicológicas" de sus elecciones; tienen una idea muy pobre sobre lo que puede ser la explicación biológica o psicológica de un fenómeno.)

[28] Para una análisis inteligente de la interacción entre la evolución genética y la evolución cultural en la ética, véase Campbell 1975.

La actividad comunitaria de la persuasión, reflexión y evaluación recíproca ¿prevalecería entonces sobre los intereses más elementales de nuestros antepasados? ¿Cuánto puede abarcar este proceso reflexivo? Hablamos a la gente para persuadirla de la bondad de determinadas acciones; también sistematizamos nuestras opiniones en principios y códigos morales; luego reflexionamos sobre estos principios en la teoría de la ética y acto seguido nos ocupamos de "metaética", en donde reflexionamos sobre la naturaleza de las teorías éticas, y así proseguimos, considerando las implicaciones que tiene, para el problema de la libertad y el libre albedrío, la semejanza con la *sphex* de nuestra capacidad de ocuparnos de la metaética, y así sucesivamente. ¿Todo esto es, por lo menos, eficaz?

Podemos pensar, en primer término, que nadie desea detener la espiral ni descubrirse esencialmente incapaz de seguir adelante. Pero, de hecho, la afirmación de que se ha agotado la capacidad humana de pensar y metapensar no es una afirmación insólita en los discursos filosóficos. En ocasiones, el vértigo intelectual se instala y comenzamos a trastabillar.

Nietzsche, por ejemplo, posee el don de hacernos tambalear cuando buscamos posiciones de ventaja más altas.

> Este problema del *valor* de la piedad y de la moral de la piedad... parece, a primera vista, algo meramente desgajado, un signo de interrogación solitario; pero aquel que se obstine en el problema y *aprenda* a formular aquí las preguntas, experimentará lo que yo he experimentado: una nueva y formidable perspectiva se abre ante sus ojos, una nueva posibilidad le sobreviene... Necesitamos una *crítica* de los valores morales, *el valor de esos mismos valores morales debe ser cuestionado.* (*Sobre la genealogía de la moral*, Prefacio.)

Foot (1973, pág. 137) comenta: "¿Pero cómo puede valer un valor? La sola idea de una cosa semejante hace que la cabeza me dé vueltas". Es posible, ciertamente, evaluar algunos de nuestros propios valores al mismo tiempo que adherimos a otros, pero Foot se confiesa mareado ante el proyecto, en apariencia, más radical de Nietzsche.

¿Es ésta, entonces, la cúspide de la pirámide? ¿O se trata más bien de la cúspide de *nuestra* pirámide, el punto donde *revelamos* nuestra semejanza con la *sphex* y donde forcejeamos tontamente, luchando contra un pensamiento más alto? Es, ciertamente, difícil y desagradable formularse a sí mismo estas y otras preguntas similares, lo cual explica, quizá, la escasez de análisis cuidadosos sobre el tema. Uno de los pocos que conozco es el ensayo

de Strawson "Freedom and Resentment" (Strawson 1962),[29] donde se plantea la controversia entre un optimista y un pesimista sobre las implicaciones del determinismo respecto de la visión de nosotros mismos como agentes o "participantes". El pesimista afirma que el determinismo tiene ciertas implicaciones deplorables qu e"podrían o deberían" conducirnos a abandonar lo que Strawson llama"la actitud participante" hacia nosotros y hacia los demás. Strawson responde:

> Suponer que esto pueda ocurrir no implica, en apariencia, contradicción alguna. De modo que deberíamos decir que no es inconcebible que ocurra. Pero me inclino fuertemente a pensar que para nosotros, tal como somos, es prácticamente inconcebible. Pienso que el compromiso humano de participar en las relaciones interpersonales comunes está profundamente arraigado y es demasiado incondicional para tomar en serio la idea de que una convicción teórica general pueda cambiar así nuestro mundo, de modo que ya no existan cosas tales como las relaciones interpersonales, según las entendemos habitualmente... (Strawson 1968, pág. 82.)

Teniendo en cuenta cómo somos, resulta prácticamente inconcebible abandonar la actitud participante. Estamos arraigados en esa posición como seguramente lo está el *sphex* en su limitada posición ventajosa. Si ésta es una ilusión, se trata entonces de una ilusión estable (e irresistible, aunque podemos anularla con un juicio). En el capítulo cinco, examinaremos el problema de si estamos atascados, y si lo estamos, qué deberíamos hacer en ese caso.

Entretanto, la conclusión de que hemos alcanzado la cúspide de *nuestra* pirámide no debería ser necesariamente una ocasión para sentirnos incómodos o menospreciarnos, como si existiese o pudiese existir un observador "superior" que fuera para nosotros lo que nosotros somos para la *sphex*. Las perspectiva fue discutida de un modo esclarecedor pero no concluyente por los filósofos (Davidson 1973, Rorty 1978 y Fodor 1983, págs 123-126). Es difícil imaginar cómo podrían convencernos en un sentido u otro. Incluso si ese ser es inconcebible para la razón, nos está presumiblemente prohibido argumentar (a la manera corriente) que si es inconcebible es imposible, pues del mismo modo somos inconcebibles para la *sphex*. Imaginar un ser superior es, después de todo, un desafío a la tradicional inferencia respecto de la imposibilidad de *nuestra* incapacidad de comprender.

[29] Véanse también las excelentes discusiones sobre este ensayo en Ayer (1980) y en Bennett (1980).

Al mismo tiempo, deseamos protegernos del oscuro tráfico de misterios, y tenemos derecho a exigir razones para suponer la existencia de una posición ventajosa tan alta, inaccesible para nosotros. No nos inducirán a imaginar un ser que sea meramente un prodigioso calculista o una enciclopedia andante, como el físico omnisciente de Laplace, ya que es improbable que el esquema conceptual de dicho ser sea superior al nuestro. Fodor señala que "uno no debería dejarse impresionar por los argumentos *a priori* cuyo objeto es probar, por ejemplo, que la verdadera ciencia *debe ser* accesible a las arañas" (Fodor 1983, pág. 126). Pero esto no excluye la posibilidad de argumentar que la verdadera ciencia debe ser accesible para nosotros, pues simplemente se limita a enrostrarnos nuestro parentesco con las arañas (y las avispas), sin evaluar lo que implican las enormes diferencias (otro abuso del generador de intuiciones).

Aun en el caso de existir posiciones más altas de ventaja, ¿nos importarían? Strawson piensa que no. Si la pregunta no se refiere a lo que haríamos en el caso de convencernos de la verdad del determinismo, sino a lo que racionalmente tendríamos que hacer, no podríamos ni deberíamos apelar a un tipo de racionalidad sobrehumana, oscuramente imaginada, afirma Strawson, sino más bien, a la racionalidad que hemos creado siendo las criaturas que somos y viviendo la vida que vivimos.

> ...Si pudiéramos imaginar lo que no tenemos —esto es, elección en esta materia—, podríamos entonces elegir racionalmente sólo a la luz de la evaluación de las ganancias y las pérdidas respecto de la vida humana, de su empobrecimiento o enriquecimiento... (Strawson 1968, pág. 84.)

La racionalidad que hemos conformado no es, por cierto, perfecta. En este capítulo he ensayado "Sólo un cuento" considerablemente optimista. Un cuento más fiel a la historia hubiese sido sin duda más oscuro. Los "Sólo cuentos" no deben pasar por alto la posibilidad del cáncer cognitivo o del cáncer conativo. Quizá fuimos diseñados para olvidar ciertas verdades o para inventar obsesivamente falsedades confortables, protegiéndonos de la irritante verdad con una perlada pátina de persuasivos mitos. ¿Aparecen las cicatrices de nuestra evolución genética y cultural? ¿Nos sorprende que lo hagan? ¿Acaso esas cicatrices nos mutilan? Como esta pregunta tiene una respuesta menos obvia y por el momento irrelevante para el problema del libre albedrío, me ocuparé de ella en los próximos capítulos.

La racionalidad que nos prodigó la naturaleza es práctica; produce cambios al movernos, la mayoría de las veces, en la direc-

ción adecuada. Pero no debemos suponer que es sólo práctica o que es un don rígidamente sometido al servicio de los fines biológicos que le dieron nacimiento. Como ya hemos visto, una parte especialmente poderosa de este don proviene de nuestra capacidad de lenguaje, que sin duda sólo podía emerger bajo la presión de exigencias y restricciones muy específicas (aunque oscuramente comprendidas) del entorno. Algo tan complejo como el lenguaje jamás hubiera podido evolucionar a menos que sirviera a una importante función biológica. Pero una vez en la escena de la evolución, el don del lenguaje abrió un espacio para que ingresaran toda suerte de intentos biológicos triviales e irrelevantes o barrocos (no funcionales): chismes, acertijos, poesía, filosofía. Viendo cómo la evolución crea la razón práctica, vemos también cómo la evolución puede dar nacimiento a la razón no práctica.

Resumiré ahora las conclusiones de este capítulo. No he suministrado la prueba de que la racionalidad sea posible para entidades físicas. ¿Cuál debería ser el objetivo de la prueba? Presumo que ya tenemos suficientes razones para creer que somos entidades físicas y que somos racionales. Pero necesitamos comprender por qué fue así y cómo pudo llevarse a cabo.[30] Las especulaciones biológicas nos ayudan a imaginar con más detalle que nuestra facultad para ser movidos por razones está oculta en el universo; al hacerlo, nos libran al menos de la tentación de retraernos dentro de dogmas absolutistas acerca de nuestra condición de agentes.

La perfecta voluntad kantiana, que sería capaz de responder con perfecta fidelidad a todas las buenas razones, es una imposibilidad física, ni el determinismo ni el indeterminismo se pudieron adecuar a ella. Pero eso no nos deja en la difícil situación de la *sphex*. No somos infinitamente sino tan solo extraordinariamente sensibles y versátiles catadores de razones. Si esto no es suficiente, si el libre albedrío y la responsabilidad exigen una *libertad absoluta* respecto de nuestra semejanza con la *sphex*, si un mínimo de responsabilidad no puede sobrevivir en nada que no sea un entorno perfectamente estéril, entonces debemos desechar por completo la visión científica de nosotros mismos o bien admitir la derrota. Pero estas terribles alternativas no nos han convencido hasta el momento. Hemos empezado a vislumbrar cuánto de lo que queremos que sea verdad puede ser esclarecido —no amenazado— por la aplicación de la visión científica, materialista.

[30] Nozick (1981) insta a los filósofos a considerar la posibilidad de abandonar la prueba formal en favor de una clase particular de explicación filosófica; esta explicación nos induciría a ver por nosotros mismos cómo aquello que deseamos creer podría ser posible. En mi opinión, se trata de un consejo excelente y he tratado de que este capítulo (y, en realidad, todo el libro) sea un ejercicio de explicación a la manera de Nozick.

3

Control y autocontrol

Nadie ha dicho jamás
que los termostatos no controlan
la temperatura
porque el determinismo es verdadero.
—Robert Nozick (1981, pág. 315)

1. "Debido a circunstancias que están fuera de nuestro control"

En las discusiones sobre el determinismo, se nos invita con frecuencia a situarnos mentalmente en un tiempo anterior a la aparición del hombre, e incluso anterior a la aparición de las avispas o de las moléculas orgánicas en el planeta. Se señala que si el determinismo es verdadero, entonces el estado de cosas en ese momento (un instante arbitrariamente elegido) ha determinado todos los acontecimientos subsiguientes del presente y del futuro. Según Laplace:

> Una inteligencia que conociera en un momento dado todas las fuerzas que actúan en la naturaleza y la situación de los seres que la componen —que fuera suficientemente vasta para someter esos datos al análisis— podría expresar en una sola fórmula los movimientos de los cuerpos más grandes del universo y los del más pequeño de los átomos. Nada resultaría incierto para ella y tanto el futuro como el pasado estarían presentes ante su mirada. (Laplace, 1820.)

Esta perspectiva nos parece aterradora, pues implica que todos nuestros actos están determinados por acontecimientos del pasado remoto, sobre los cuales no tenemos ningún control, y que, por lo tanto, nunca controlamos realmente nuestros actos. Estamos

controlados por el pasado o por acontecimientos del presente causados por acontecimientos del pasado que no controlamos.[1] En el mejor de los casos, gozamos de un control ilusorio. Pero, en realidad, estamos completamente a merced de factores externos, encerrados en un cuento que se escribió en el alba de la creación, como un títere que estuviera destinado a serlo aun antes de que se tallaran y pintaran sus rostros de madera. Rara vez se examina en detalle esta explicación del temor que inspira el determinismo, lo que no nos sorprende, pues está tan plagada de confusiones que no resiste la luz del día.[2]

Lo que nos provoca temor en el determinismo es, en gran medida, la perspectiva de que pueda abolir el control.[3] Y, definitivamente, no deseamos perderlo ni ser controlados por algo o alguien como si fuéramos títeres o marionetas.[4] Queremos *controlarnos* y controlar nuestro destino... Pero ¿qué es el control y cuál su relación con la causalidad y el determinismo? Curiosamente, los filósofos raras veces han formulado esta pregunta, a pesar de su innegable importancia.[5]

[1] Cualquier proposición verdadera acerca de cómo eran las cosas antes de que existieran seres racionales es tal que nadie puede (ni pudo jamás) determinar si es verdadera (Van Inwagen, 1983; pág. 149).

[2] Muchos autores han manifestado sus sospechas respecto de la táctica de remitirse a un momento prehistórico. Nozick (1981) considera "cómo tendemos naturalmente a exponer nuestra preocupación por el determinismo":

> ¿Por qué es tan natural hacer hincapié en las causas anteriores a nuestro nacimiento?; o mejor, ¿qué función cumple? o ¿Qué otras posibilidades se impugnan? Es evidente que si las causas se retrotraen a un tiempo anterior a nuestro nacimiento, no las controlamos, y puesto que ellas controlan nuestras decisiones, nosotros no controlamos nuestros actos (pág. 314).

Y las sospechas de Slote se despiertan porque "con la única excepción de Wiggins, quienes proponen la nueva versión del argumento de la incompatibilidad señalan, en particular, los acontecimientos ocurridos antes del nacimiento del agente cuando se refieren a las cosas que el agente no puede controlar" (Slote 1982, pág. 19).

[3] Fischer (1982) señala que "comúnmente se piensa que el determinismo erosiona el control". Lucas (1970) afirma, sin basarse en ningún argumento, que "el determinismo priva [a las personas] de toda verdadera decisión acerca de sus vidas" (pág. 28).

[4] "Es necesario señalar que no somos dispositivos extremadamente simples de *input-output*, pues se efectúan muchos procesamientos internos que comportan circuitos de retroalimentación y de *software*; quizás ello nos hace títeres más complicados pero títeres al fin. Ciertamente, muchas de estas cosas suceden 'dentro' de nosotros. ¿Qué es preferible, ser un muñeco manejado por una mano o un muñeco manejado por cuerdas?" (Nozick, 1981, pág. 310).

[5] Hasta donde sé, Greenspan (1978) ha hecho observaciones tangenciales sobre el concepto de control, al igual que Holmstrom (1977), Fischer (1982), Nozick (1981) y Slote (1982); ninguno de ellos se ha referido a ninguna literatura anterior sobre el tema.

Vamos a poner en práctica, entonces, un poco de "filosofía del lenguaje ordinario" (tan largamente postergada), a fin de averiguar qué pensamos en realidad cuando anhelamos el control y tememos su pérdida.[6]

2. Control simple y autocontrol simple

¿Qué significa el control? ¿Qué significa controlar algo? Luego de responder a esta simple pregunta, nos ocuparemos del control de nuestro destino, de las diversas clases de autocontrol y de algunos detalles más sutiles de este concepto básico. Los diccionarios definen el término "controlar" acumulando sinónimos: *gobernar, dirigir, dominar, verificar, subrayar,* signo evidente de que el concepto es demasiado fundamental para admitir una definición directa.

La idea básica de control, elevada a la categoría de concepto técnico en la cibernética y en la teoría de los autómatas, es (en términos ordinarios) la siguiente: A *controla* a B si, y sólo si, la relación entre A y B es tal que A puede *inducir* a B a todos los estados normales en los que A *desea* que se encuentre B. (Si B es capaz de encontrarse en un estado *e* y A desea que B se encuentre en ese estado este estado *e* pero no puede lograrlo, el deseo de A se frustra y, en ese caso, A no controla a B.) Esta definición pone de manifiesto que lo controlado debe poseer diferentes estados en los cuales se pueda hallar,[7] y que los estados de quien controla deben incluir deseos (o algo semejante al deseo) referidos a los estados del otro. (En el capítulo 1 mencionamos el tumor cerebral que no podría

[6] Este ejercicio se relaciona con las primeras defensas "en lenguaje ordinario" del libre albedrío contra el determinismo; en particular, la apelación al tipo de argumento del Caso Paradigmático, quizá mejor ejemplificado por Flew (1955, véanse especialmente las págs. 149-163. Véanse también van Inwagen, 1975 y 1983, sobre la crítica del argumento de Flew). A diferencia de Flew, no afirmo que contamos con "criterios" comunes sólidos para determinar si tenemos el control de algo que sea a prueba de cualquier descubrimiento científico. Sin embargo, como Flew, subrayo un hecho empírico obvio: nos preocupamos por el concepto de control porque nuestro *concepto ordinario de control* nos permite distinguir las circunstancias que nos agradan ("tener control", "controlarse a sí mismo", etc.). Un concepto que no hiciera tales distinciones no nos permitiría saber si debemos preocuparnos por tener el control, en *ese* sentido. Puesto que el problema del libre albedrío es un problema que nos preocupa "esencialmente", deberíamos *comenzar* toda investigación sobre el tema a partir del supuesto tentativo y refutable de que, cualquiera sea la versión que existe de concepto, es una de las versiones "importantes".

[7] Nótese que este "pueda" es la rana de Austin en el fondo del jarro de cerveza, tema que retomaremos en el capítulo 6.

controlar los estados del pobre Jones a menos que fuera suficientemente "astuto" y tuviese deseos relativos a los estados de Jones.)

En el siguiente ejemplo, veremos algunos aspectos del concepto ordinario de control. Supongamos que usted ha comprado un avión de juguete que funciona a motor. Lo pone en marcha, el avión sale volando y al cabo se estrella contra un árbol. "Necesita control", piensa usted; acto seguido, le ajusta un cable largo a la punta del ala. Ahora está en condiciones de subirlo a voluntad, mientras el avión gira y gira en círculos. "No hay suficientes *grados de libertad*", piensa usted; de modo que reemplaza el cable por un sistema de "control remoto" y puede controlar entonces la dirección, la altura, la velocidad, los giros, las picadas y la inclinación del aeroplano. Existen más grados de libertad que sólo estarán bajo su control cuando descubra los parámetros de los grados de libertad del aeroplano y las relaciones causales entre esos parámetros y los movimientos de su mano en el control remoto (un problema epistémico nada trivial, que deberá resolverse antes de ejecutar el control).

Si el aeroplano escapa a la zona de influencia de su sistema de control remoto, volará *fuera de control* pero estará sometido, como de costumbre, a las fuerzas causales que actúan en las inmediaciones. Quizá vuelva a ingresar al área de control y entonces hará lo que usted quiera. Hasta cierto punto, naturalmente. Un aeroplano no puede permanecer inmóvil en el mismo sitio o ascender verticalmente, pues esas propiedades pertenecen a los helicópteros, no a los aviones. Sin embargo, a diferencia de los aviones, los helicópteros no vuelan a grandes velocidades, ni hacen lupines, ni giran sobre sus eje longitudinal. Se pueden controlar los estados o actividades de una cosa siempre que pertenezcan a la clase de actividades que son capaces de realizar. Cuando nos referimos a lo que *puede hacer* un avión de juguete o a *sus grados de libertad*, no le estamos atribuyendo el tipo de acción o de libertad que le atribuimos a un agente moral. Sin embargo, estas facultades más simples son los requisitos indispensables para una acción y una libertad superiores. Antes de sacar conclusiones metafísicas apresuradas con respecto a las limitaciones de estos conceptos más simples, debemos observar su funcionamiento con mayor cuidado. ¿Cómo sabe usted con certeza cuándo tiene el control del aeroplano? Supongamos que lo acompaña una persona que tiene una caja de control igual a la suya. ¿Cuál de los dos —se pregunta usted— controla el avión? ¿El aeroplano todavía responde a su voluntad? Si no lo hace, usted seguramente ya no lo controla. ¿Qué ocurre si otra persona imita sus movimientos? En ese caso, ¿quién controla el avión? Existe la posibilidad de que su control sea ilusorio,

pero se trata de una posibilidad que puede investigarse empíricamente.[8] Usted está en condiciones de realizar un clásico experimento "controlado". Utilizando lo que a menudo se denomina el método de concordancia y diferencia de Mill, puede modificar deliberadamente las circunstancias y anular la relación entre el movimiento de su mano y el de la mano del imitador, identificando así los factores que intervienen realmente en el control del avión.

Supongamos que la persona que lo acompaña controla el aeroplano. Si usted es capaz de controlar la mímica del otro, es decir, si logra que el otro persista en su "estado de imitación", podrá controlar su aeroplano por medio de la otra persona. Se aducirá que este método es menos seguro que el de manipular directamente el control remoto (que, a su vez, es causalmente menos directo que el control manual). Pero el control que se ejerce por medio de otro agente es bastante seguro para el capitán de un barco, por ejemplo, que rara vez toca el timón. (Como veremos más adelante, hay muchas maneras de controlar algo, controlando a un agente intermedio.)

Hay hechos obvios que vale la pena mencionar antes de entrar en asuntos más complicados. Cuando usted controla su avión no lo hace controlando todas las causas que lo influyen. El clima, la densidad del aire y la fuerza de gravedad, por ejemplo, están fuera de control, y son las fuerzas principales que actúan sobre el aeroplano. El hecho de que el avión esté constantemente sometido a la ley de la gravedad no le impide controlarlo, pues, en cierta medida, la gravedad ayuda al control. Siempre que usted *conozca* sus efectos sobre el aeroplano. Pero una súbita pero inesperada corriente de aire puede desbaratar transitoria o permanentemente dicho control. La diferencia entre la gravedad y las corrientes de aire son (relativamente) inesperadas. Pero si usted es lo bastante inteligente para preverlas, colaborará en el control del aeroplano en la misma medida en que colabora la ley de gravedad. Los marinos se valen de este conocimiento (los efectos de las corrientes de aire se perciben rápidamente en el agua) para controlar la dirección y la velocidad de sus embarcaciones. Un perro que tira de un trineo por un camino que le es familiar tendrá más control —gracias al conocimiento previo de las condiciones del camino— que un perro que emprende ese trayecto por primera vez.

[8] En principio, la ilusión es difícil de detectar, como lo ha confirmado el pobre Dennett al final de "Where Am I?", cuando su *output* cerebral no pudo conseguir que su cuerpo se moviera y éste se desparramó torpemente por el aire, mientras el cerebro de un ordenador ejercía realmente el control. Un ejemplo más lejano en el tiempo lo constituye el genio maligno de Descartes, que bien pudo haberlo engañado con respecto a todas las cosas.

El conocimiento previo es, pues, lo que hace posible el control. Si conocemos a tiempo las circunstancias que obviamente están fuera de nuestro control, éstas no impedirán ni anularán el control que deseamos ejercer sobre ciertos acontecimientos; conocerlas a tiempo significa planear la acción de acuerdo con nuestras expectativas y tomar recaudos para prevenir, evitar, defendernos, adecuarnos y explotar tales circunstancias. Solamente estamos *a merced de* las cadenas causales que nos sobrevienen sorpresivamente o que no nos dejan opción alguna para evitar los efectos indeseables (en el capítulo 5 analizaremos el concepto de "evitar"). Boston se encuentra a 3.000 millas de Oxford, una circunstancia que está por completo fuera de mi control. Pero el hecho de saberlo me permitió tomar las medidas adecuadas para llegar a Oxford y dar mis conferencias a tiempo. Como no pude modificar esa circunstancia, me acomodé a ella. En abril de 1983, previendo que habría una larga cola para la reserva de vuelos transatlánticos —una circunstancia igualmente fuera de *mi* control, pero no del control del presidente Reagan— reservé mi pasaje con suficiente antelación.

Cuando se controla algo, hay que estar en contacto con ese algo. En el caso del avión de juguete, la esfera de control está determinada por la potencia del equipo de radio (hasta cierto punto). Los ingenieros y científicos de la NASA responsables de las naves exploradoras enviadas a Marte y Venus dirán que la velocidad de la luz se transforma en un factor nada desdeñable cuando aumentan las distancias. Fue absolutamente imposible controlar en detalle este tipo de naves desde la tierra, ya que el tiempo requerido para enviar y recibir una señal era mayor que el tiempo disponible para ejecutar la acción adecuada. Las personas que desde la tierra controlaban las naves no tenían tiempo suficiente para conocer las condiciones relevantes y poder así elegir las instrucciones atinadas. En síntesis, las naves ya no podían ser *títeres*, tenían que convertirse en *robots* y controlarse a sí mismas. Por lo tanto, debían estar equipadas con "deseos" (rudimentarios) y con suficiente"conocimiento" acerca de sus circunstancias, de modo de realizar el trabajo tan bien como lo habrían hecho quienes las controlaban desde la tierra, en caso de haber estado suficientemente cerca para intervenir.[9]

Las personas que permanecían en tierra ejercían un control secundario, parcial, sobre los robots, enviando mensajes generales que indicaban cuál era el siguiente proyecto a realizar. Además, los robots habían sido diseñados para "deliberar" con ellas en ocasiones

[9] Un caso particularmente fascinante es "El Viking Autónomo" (Hutchins, 1983); y los ingenieros de la NASA resolvieron dejar de controlar el cohete espacial Viking y lo "automatizaron" cuando ya estaba girando alrededor de Marte.

especiales.[10] De modo que estos robots no estaban enteramente "fuera de control", sino que se hallaban controlados, hasta cierto punto, desde la Tierra.[11]

Como vimos, del hecho de que *x* controla a *y* no se sigue que *z* no controle también a *y* controlando a *x*. En ciertas condiciones, el control puede tener carácter transitivo.[12] Esto es verdadero incluso cuando *x* e *y* son idénticos (o sea, cuando *x* se controla a sí mismo). Puedo controlar algo que se controla a sí mismo, si puedo controlar los estados del mundo que, al controlarse a sí mismo, lo llevan a actuar. Supongamos que se reemplaza a los operadores que controlan una refinería por un sistema de control computarizado. El sistema controla la refinería tan bien como lo hacían los operadores. Observando la refinería en su conjunto —incluyendo su sistema de control— podemos decir ahora que la refinería se *controla a sí misma*. Sin embargo, si cambiamos los estados de control del ordenador, provocaremos cambios en los estados de la refinería. Por ejemplo, si modificamos la cantidad de crudo que ingresa, el ordenador tendrá que responder modificando las operaciones de la refinería según ese dato. Y, si es sensible a las fluctuaciones del precio del combustible en el mercado libre, y somos capaces de subir o bajar artificialmente el precio, controlaremos indirectamente el control del ordenador de la refinería.

Cabría afirmar aquí que el tipo de control que podríamos imponer al ordenador es muy distinto que el que podríamos ejercer (aplicando las mismas tácticas) sobre los operadores que controlaban la refinería antes de la automatización. Pero no existe ninguna evidencia de que esto sea verdadero. Los seres humanos pueden tener más recursos que el ordenador imaginario para eludir el control, pero, como no se han especificado los medios que podría utilizar el ordenador para evitar el control, no nos asisten razones

[10] Dawkins (1976) describe cómo los genes "controlan indirectamente la conducta de sus máquinas de supervivencia, no tirando directamente de las cuerdas, como en el caso de las marionetas, sino indirectamente". Para ilustrar cómo un agente o cuasiagente puede *delegar* el control en otros agentes que no están dentro del área de influencia directa del primero, utiliza el excelente cuento de ciencia-ficción: *A for Andromeda*, de Hoyle y Elliot (Dawkins, págs. 56-57, reimpreso en Hofstadter y Dennett, 1981).

[11] Sería útil reflexionar sobre la responsabilidad que les cabe a las personas de la Tierra por los "actos" de sus robots. El jurado querrá saber si el robot estaba fuera de control o si había dejado de controlarse a sí mismo, etc. ¿Logró *ejercer* verdaderamente el control sobre sí mismo? ¿Hasta qué punto se autocontrolaba? Como ya he dicho, el autocontrol que poseen los robots no es el mismo que poseemos *nosotros*; no es el tipo de autocontrol que uno va aprendiendo a medida que se aproxima a la adultez. El tema se retomará en el capítulo 4.

[12] Nozick (1981, pág. 314) formula la pregunta: "¿pero la noción de control es transitiva?", y no se molesta en responderla.

para sustentar esta creencia. Quizás el ordenador posee un tipo de automonitoreo encubierto, versátil y engañoso que le permite ser mucho menos controlable que un operador.

Además, aunque fuera cierto que en tales circunstancias las personas están menos sujetas a control, no podemos negar que el ejercicio de controlar a quienes controlan es, precisamente, una de las actividades que nosotros, los seres humanos, repudiamos y tratamos de evitar. No nos gusta ser controlados de esa forma por otras personas, por lo tanto, nos sentimos efectivamente controlados (hasta cierto punto) por esa clase de actividades. No nos sentimos inmunes al tipo de control al cual es vulnerable el ordenador; nuestra humanidad y nuestra facultad de elegir libremente no nos eximen de ser objetos de este tipo de control (véase Frankfurt, 1973).

3. Control sin agente y el concepto de causación

Cuando un agente intenta controlar a otro agente (que, a su vez, puede estar tratando de controlar al primero), las cosas se complican. Pero si deseamos avanzar en la comprensión de este fenómeno importante, es necesario despejar, en primer término, la confusión que se halla en el fundamento mismo de nuestro concepto de control. Como acabamos de ver, un agente externo puede controlar el sistema de control de la refinería controlando los elementos del entorno —por ejemplo, bajando el precio del combustible en el mercado—. Pero si nadie participa en esta manipulación, ¿debemos inferir que el mercado o sus fluctuaciones controlan al controlador de la refinería? ¿Diremos que el entorno es una especie de agente mínimo que *ejerce* el control?

Esta visión de las cosas no nos es desconocida. Constituye un tema central de la teoría del conductismo de B.F. Skinner (véase Skinner, 1953, en especial "The Controlling Environment", págs. 129-141 y "Self-Control", págs. 227-242). Esta concepción nos resulta en extremo desalentadora, pues, entre otras razones, sugiere subliminalmente la oscura idea de que el entorno *quiere* que hagamos algo determinado y lo *consigue* actuando sobre el deseo. Pero no es así como Skinner ve las cosas. Según su teoría, no existen fenómenos tales como los deseos, ni siquiera en las personas o en otros organismos; por lo tanto, el entorno carece de deseos o de intenciones, ya sean buenos, malos o indiferentes. No obstante, distingue la mera causalidad ambiental (por ejemplo, los efectos de sol, el viento y la lluvia sobre la piel de un determinado organismo) del control ejercido por esos elementos del entorno tomados como estímulos. A diferencia de nuestra desalentadora visión, para Skinner se trata de un control benéfico —tal como lo desearíamos,

si existiese esa cosa llamada deseo—. "Podemos estar en desacuerdo respecto del grado o la naturaleza del control que [el mundo circundante] ejerce sobre nosotros, pero es evidente que existe algún control. La conducta debe adecuarse a las circunstancias" (Skinner 1953, pág. 129). Por lo tanto, Skinner diría que los ingenieros de la NASA tienen como objetivo diseñar naves exploradoras, de manera que puedan ser adecuadamente controladas por el mundo marciano que las circunda.

Skinner habla constantemente de la conducta que está "bajo el control" de diversos estímulos, pero no enfatiza la diferencia entre ser adecuadamente controlado por el mundo circundante y serlo inadecuadamente. Si un robot que efectúa exploraciones en Marte fuera inducido por las pautas del entorno a destruirse a sí mismo, estaría actuando bajo el control skinneriano de los estímulos de Marte, pero le "gustaría" destruirse (si es que algo puede gustarle). Si, como sugiere Skinner,[13] ser controlado por estímulos es algo bueno, lo será sólo cuando los estímulos nos controlen de manera correcta.[14]

¿Cuál es la manera correcta? Obviamente, *la que desaríamos si lo supiéramos todo acerca del mundo y de nuestros intereses.* Así pues, la manera correcta es la veraz o verídica. Si el mundo nos induce inexorablemente a conocer la verdad sobre sí mismo, nos sentiremos bastante felices en la medida en que el mundo actúe sobre nosotros. Como dijo Wiggins: "El libertario... debería estar satisfecho de permitir que el mundo le dictara al hombre libre cómo *es* el mundo. La libertad no consiste en el ejercicio, plausible pero irrelevante, del derecho a volverse loco sin que la realidad interfiera o distraiga" (Wiggins 1973, pág. 34). Ojalá nuestras creencias estuvieran controladas por la manera en que las cosas realmente son (véase la definición de conocimiento como "la creencia que rastrea la verdad", en Nozick 1981, capítulo 3).

[13] Por un lado, siempre resultó chocante y teóricamente contradictorio que Skinner suprimiera la diferencia entre el *mero* control bruto ejercido por el entorno y la parte más agradable, el control bruto *sistemáticamente útil* ejercido por el entorno. Por el otro, tiene sus razones: como el adaptacionista, ha visto que ese control *tiende* a ser benéfico. (La clase de control sistemáticamente asesino que imaginamos en párrafos anteriores y que conduce a la destrucción del robot, se autocancela a corto plazo, aniquilando a su presa.) Pero Skinner sobrestima su punto de vista y provoca merecidas suspicacias. Por cierto, no todo "control ejercido por estímulos" es bueno simplemente por ser un control y Skinner está lejos de convencernos de que es así.

[14] Compárese con la confianza en la "manera correcta" que aparece en las teorías causales del conocimiento y de la referencia, donde la cadena causal entre el objeto y el pensamiento debe ser del tipo "correcto" —como de costumbre, la naturaleza de "lo correcto" será especificada en otra oportunidad—.

Un mundo verdaderamente benévolo nos mantendría informados acerca de sí mismo; por supuesto, ¡las noticias tendrían que ser buenas! No esperamos un lecho de rosas, pero al menos un mundo benévolo que nos dicte "lo que hay que hacer"; por el contrario, un mundo malévolo nos "engañaría", induciéndonos a "caer en trampas" y a "apartarnos del camino", con el propósito de frustrar nuestros planes. El mundo, ciertamente, no está ni "a favor" ni "en contra de" nosotros, aunque algunas de sus partes nos resulten tonificantes o tóxicas, amigables u hostiles.

La tendencia benéfica que Skinner atribuye —de manera muy poco plausible— al entorno inmediato de "los estímulos" presenta, en realidad, dos niveles: una tendencia benéfica del entorno, en el largo plazo evolutivo, a diseñar criaturas que poseen *en sí mismas* una tendencia benéfica a efectuar discriminaciones justas (para ellas mismas). El entorno no está diseñado para *decirnos* lo que debemos hacer; somos nosotros los que estamos diseñados para *resolver* lo que tenemos que hacer a partir del entorno indiferente. De manera que es imprudente afirmar que el entorno nos controla,[15] aun a la luz de esta tendencia indirecta pero sistemáticamente benéfica a largo plazo.

Como vimos, el concepto de control de Skinner es muy diferente de nuestra noción ordinaria de control, pues en dicho concepto no está implicado el agente o quien alberga los deseos. Decimos que A controla a B, en el sentido skinneriano, si los cambios de A son fielmente registrados o reflejados en los cambios de B. No se trata de un concepto excéntrico, aunque difiera de nuestra noción corriente. En el mismo sentido, los surcos del disco controlan las vibraciones de los parlantes, y la cantidad de lluvia caída en un determinado período controla la altura del río. Lo que pone en funcionamiento a A no necesita ser un agente y lo que hace funcionar a B no necesita ser un organismo o algo que tenga intereses. Por lo tanto, no importa si los cambios de B controlados por los cambios de A son adecuados para el organismo que es B (o en el cual reside B). Podría suponerse que el concepto de Skinner es el concepto común de causación física con otro nombre, pero en realidad ocupa una zona intermedia y fomenta el desplazamiento ilusorio de la causación hacia el control que alimenta nuestro miedo al determinismo.[16]

[15] En realidad, Monod encuentra razones para insistir en lo contrario: el entorno no controla ni puede controlar a un ser viviente. Un ser viviente es entonces una estructura que da pruebas de poseer un determinismo autónomo, preciso, riguroso, que implica virtualmente una "total" libertad respecto de los agentes o a las condiciones externas que pueden, sin duda, obstaculizar este desarrollo pero de ninguna manera gobernarlo o guiarlo; en todo caso, no imponiendo un esquema de organización al objeto viviente (Monod 1971, págs. 10-11).

Parte de la explicación de la confusión entre causación y control proviene de nuestras prácticas científicas comunes y, en especial, de nuestra manera de entenderlas. En un experimento típico, el investigador *controla* cuidadosamente las diferentes condiciones con el propósito de establecer —como lo hizo quien controlaba a distancia el aeroplano del ejemplo— que una variable particular, la variable dependiente, se encuentra *bajo el control del investigador* (en el sentido corriente) al estar vinculada con ("controlada por", diría Skinner) una variable independiente o con estímulos que el investigador controla. El fundamento de esta práctica es conocido. Se observa una pauta en un determinado fenómeno y se pregunta cuál es la causa. Si el investigador puede dirigir la pauta "a voluntad", las causas son descubiertas (esto es, las causas implicadas en la cadena de control que va del investigador al fenómeno observado). Para lograr estos efectos, debemos lograr que las circunstancias sean epistémicamente manejables, es decir, debemos aislar el fenómeno de las influencias externas "impredecibles" y hacer que el entorno inmediato sea lo más simple y uniforme posible a fin de rastrear los elementos relevantes y, por consiguiente, lograr efectos repetibles a voluntad. Cuando pensamos en la causación, en general, influyen en nuestra imaginación los éxitos obtenidos bajo estas "condiciones controladas".

Sabemos perfectamente que la causación puede ser inescrutable —perdida por completo en una maraña inextricable de coincidencias— y ser todavía causación. ¿Qué hizo posible que *este* grano de arena se encuentre exactamente en este lugar de la playa? Algo lo hizo, o mejor, millones de cosas en el transcurso de millones de años lo hizo posible. Generalmente pasamos por alto este tipo de cosas. Pero si el grano de arena pertenece a una huella, entonces nos ocuparemos de una sola causa: la causa de su hundimiento respecto de la posición en la que presumiblemente se encontraba. Cuando pensamos en casos de causación, pensamos casi invariablemente en aquellos en los cuales las relaciones son manifiestas, en donde es manifiesta la realidad o al menos la probabilidad del *control por medio de un agente*. Después de todo, hablamos de la "causa" de un fenómeno o acontecimiento sabiendo que la causación presenta siempre una infinidad de factores de los cuales podemos extraer, a veces, mediante un acto intelectual, un determinado elemento relevante —es decir, epistémicamente manejable y, por añadidura, controlable—, "la" causa. Estos son los "casos buenos y evidentes"

[16] El concepto de control de Skinner se asemeja al concepto de señal fiable (no ruidosa) en la teoría de la información. La variación en B lleva información sobre la variación en A. Véase Dretske, 1981.

de causación. Tendemos a olvidar los casos igualmente buenos que son virtualmente indescriptibles y totalmente incontrolables por nosotros.

En efecto, algunos de esos casos de causación se denominan "aleatorios" (tirar dados, barajar naipes, arrojar la moneda) precisamente porque son incontrolables. Podemos pensar que los saltos de la bola de la ruleta están completamente determinados por el estado de la rueda y por el impulso inicial dado a la rueda y a la bola. Pero, como este comportamiento ejemplifica (y fue diseñado para ejemplificar) un proceso incontrolable ("aleatorio"), es fácil olvidar que no está más exento de causación que la bola de billar que rueda inexorable por la mesa hacia un sitio deliciosamente predecible gracias al control de un jugador avezado —que planeó el tiro confiando en la uniformidad de la mesa—. Un experto en tirar la bola de la ruleta o en barajar los naipes que pudiera controlar el resultado como lo hace el buen jugador de billar, eliminaría el factor de "azar" en estos juegos y se le prohibiría la entrada a los casinos, ya que se da por sentado que en dichos juegos el resultado debe ser independiente de los deseos y creencias de sus participantes.

La inteligencia superior de Laplace probablemente no notaría la diferencia, y si le ofrecieran un trabajo en un casino podría controlar el juego de la ruleta así como el buen jugador de billar puede controlar su juego.[17] Sin lugar a dudas, una inteligencia tan sobrehumana podría controlarnos tan fácilmente como nosotros controlamos a la *sphex*, pero este ser superior es sólo uno de los tantos agentes imaginarios de los que no necesitamos preocuparnos. Quizás estos agentes habitan en otros planetas. En ese caso, sería razonable tenerles miedo. Lo que no es razonable es evocarlos a partir de la premisa del determinismo.

4. Agentes en competencia

En el transcurso de la evolución, el entorno ha realizado la brillante tarea de diseñarnos; no obstante ello, no nos controla, pues

[17] No es necesario suponer que la inteligencia superior tiene conocimiento total de sus propios microestados y transiciones (algo lógicamente imposible), pues sus interferencias con el mundo (sus actos en el casino) podrían ingresar a través de ciertos filtros cuyas propiedades conociera íntimamente —se puede aplicar el mismo principio a un ser humano si en la mesa de juego se le permitiese utilizar una máquina de tirar dados construida con precisión—. Ante la imposibilidad de un autoconocimiento completo y de sus implicaciones respecto del libre albedrío, véanse los clásicos trabajos de Popper (1951) y MacKay (1960).

no es un agente. Podría parecer que estoy promoviendo aquí una política contradictoria respecto del "uso de la imaginación". De un lado, me opongo a la personificación espuria del entorno en un rival malévolo. Del otro lado, considero que la personificación de la evolución y de sus tendencias en la sabia, benévola y vieja Madre Naturaleza, cuya "mano invisible" nos diseñó a todos, es una imagen gráfica útil. ¿No es éste, acaso, el sabio consejo de Polyana de mirar sólo el lado bueno de las cosas? Creo que no.

El antídoto estándar contra el abuso de la imagen de la Madre Naturaleza consiste en tener siempre presente que los procesos evolutivos actúan *sin prémeditación y sin metas*. Si recordamos esto cuidadosamente, podemos seguir usando la metáfora de la agencia —o la descripción "desde la actitud intencional", como la denomina en Dennett 1971 y 1978a— para describir y explicar el modelo de desarrollo que da cuenta de los buenos diseños que observamos en la naturaleza. Ahora bien, los genios del pesimismo pueden descubrir modelos antidarwinianos de malignidad sin motivo los cuales nos permitirían reformular la idea de la naturaleza como una suerte de lucha maniquea entre la Madre Maturaleza y el Mal; pero, hasta donde yo sé, nadie los ha tomado seriamente en cuenta.[18]

En ocasiones, las cosas nos resultan adversas; el entorno es indiferente, pero preferiríamos que el mundo fuese de cierta manera y no de otra. Y esto por buenas razones: ciertas disposiciones de los objetos favorecen nuestros intereses más que otras. La naturaleza "se confabula" a veces para frustrarnos, limitando nuestras elecciones y urgiendo —si no coaccionando— nuestros actos.[19] Pero si aplicamos el antídoto de los evolucionistas contra las tormentas

[18] El humor es otra cosa. Recomiendo al respecto la clásica definición de Paul Jennings en "Report of Resistencialism" donde hace referencia a la filosofía de Pierre-Marie Ventre, autor de *Resistencialisme* y de la pieza teatral *Puits clos. Les choses sont contre nous* [Las cosas están contra nosotros] es el corazón aforístico de esta teoría. Salvo sus precursores alemanes Freidegg y Hedansiecker, todos los pensadores anteriores, desde los eleáticos hasta Marx, han concedido al menos cierta legitimidad al pensamiento y esfuerzo humanos. En la cosmología resistencialista, que actualmente hace furor en París, Ventre nos ofrece una grandiosa visión del universo como Lo Uno (la cosa última, *la dernière chose*) que está "contra nosotros". El resistencialismo tiene una base científica. Todo comenzó cuando Clark Trimble, un psicólogo de Cambridge "que no era originariamente un físico" descubrió accidentalmente, "la gradual hostilidad de las cosas". La obra de Ventre nos acerca mucho más al objetivo del resistencialismo, evocado en estas palabras: "todas las cosas están fuera de control". La primera versión de "Report on resistencialism" apareció en *Spectator* y una versión revisada, en *Town and Country*, en los Estados Unidos; se ha vuelto a editar en Jennings 1963.

[19] Greenspan (1978, pág. 235) afirma que la noción de compulsión es más amplia que la de coerción: "Sólo la última requiere la intervención directa de otro agente como autor de la amenaza".

"maléficas", los acantilados "prohibidos" y otras atractivas metáforas del género, quizá seamos capaces de mantener a raya nuestra imaginación. Recuérdese que los ogros no están en el mundo sino en los ejemplos de los filósofos.

Si se le informa al piloto de una compañía aérea que hay una peligrosa tormenta en la ruta y que si entra en la zona de tormenta *puede perder* el control del avión, el piloto desviará el rumbo de la nave para preservar o aumentar el margen de control. (Esta táctica esencial de disminuir el grado de incertidumbre es un concepto profundamente esclarecedor en la teoría de la evolución. Véase Winsatt 1980.) Solamente una acción instantánea y certera preservará la marcha del avión en medio de una tormenta. Cuando la visibilidad es mayor, mayor es el margen para subsanar errores. Pero cuanto más pequeño es el margen, menos libertad de acción tiene el piloto y menos posibilidades de llegar a destino. Consciente del hecho, el piloto no sólo lucha por controlar el avión, sino que también planifica el control en el metanivel tomando medidas para mejorar su posición y evitar las circunstancias en las que se vería forzado (dados sus objetivos) a pasar por el ojo de la aguja entre algún Escila y Caribdis.

Este es un caso particularmente claro de la tan deseada libertad de acción. Deseamos un amplio margen de error y opciones abiertas que aumenten la probabilidad de controlar nuestras operaciones en cualquier caso.[20] Cuando miramos el futuro para determinar qué obstáculos seremos capaces de afrontar, deberíamos incluir por cierto nuestras limitaciones y debilidades. Si alguien adora los dulces, por ejemplo, puede adoptar la drástica medida de no tener dulces en la casa, donde es difícil no caer en la tentación de devorarlos. Las estrategias de orden superior diseñadas para maximizar la libertad de acción dependen rigurosamente del conocimiento que el estratega tiene de sí mismo.

Cuando prevemos las limitaciones futuras de nuestras opciones, la presencia o ausencia de un agente rival es verdaderamente importante, ya que un agente que recoge información e impide la retroalimentaciôm desde el exterior está en condiciones de predecir nuestras actividades y, por lo tanto, de frustrarlas sistemática-

[20] Fischer dice: "Cuando relacionamos la responsabilidad con el control, queremos decir que una persona es responsable de un acontecimiento particular sólo si existe una secuencia alternativa abierta al agente, en la cual éste pueda actuar de modo diferente (o bien producir un acontecimiento diferente al realizar un acto originado en ese acontecimiento" (Fischer 1982, pág. 34). En el capítulo cinco veremos qué significa "abierta al agente". Fischer continúa hablando del "control" —"la libertad de actuar de otra manera" (pág. 37)— pero comete aquí un error fatal, como veremos en el capítulo seis.

mente.[21] Es importante que el piloto de la línea aérea no necesite ocultar sus verdaderas intenciones a la tormenta, pues ésta podría (en caso de saberlo) cambiar el sentido de la marcha y correr detrás del avión hasta darle alcance.

Si el agente A ha de controlar el objeto B, deberá resolver el problema epistémico de identificar los parámetros de la operación de B. Cuando dos agentes compiten tratando de controlarse mutuamente, existe una competencia por la información, y cada uno de ellos se esfuerza por ocultar los propios planes y el conocimiento que posee sobre los planes del otro, etc., mientras obtiene tanta "inteligencia" como le sea posible respecto de su rival.[22] No es extraño que en tan delicadas y aterradoras circunstancias el agente reflexione: "Si estoy en desventaja, mi rival me engañará como el biólogo engañó a la pobre avispa y se las ingeniará para controlar mis actividades de un *modo equivocado*, contrario a mis intereses".

Este miedo se halla en los fundamentos mismos de innumerables generadores filosóficos de intuición.

> Considérese el caso frecuentemente debatido del neurólogo demoníaco que manipula directamente el cerebro de su víctima, induciendo todos sus deseos, creencias y decisiones. (Fischer 1982, pág. 37.)

[21] Esto es muy similar a uno de los fundamentos de la teoría de los juegos. Como señalan von Neumann y Morgenstern en el primer capítulo de *Theory of Games and Economic Behavior* (1947), en tanto un agente solitario del tipo "Robinson Crusoe" puede considerar todos los problemas como una busca de lo máximo, tan pronto aparezcan en el entorno otros agentes que busquen lo máximo, necesitará métodos radicalmente distintos de análisis, pues

> un principio que guía no puede ser formulado por el requerimiento de maximizar dos (o más) funciones a la vez... Se estaría en un error si se cree que puede ser obviada... recurriendo simplemente a los dispositivos de la teoría de la probabilidad. Cada participante puede determinar las variables que describen sus propios actos, pero no los de los otros. Sin embargo, estas variables "extrañas" pueden ser descritas, en su opinión, por supuestos estadísticos. Esto es así porque los otros agentes son guiados (como él mismo) por principios racionales —sea ello lo que fuere—. Y ningún *modus procedendi* será correcto si no intenta comprender esos principios y la interacción de los intereses en conflicto de todos los participantes (pág. 11).

[22] En el público en general existe una considerable ansiedad respecto del poder que los psicólogos pueden llegar a tener —o ya tienen— sobre la gente común, en virtud de su mayor conocimiento de la mente humana (al menos, así lo proclaman ellos). Este miedo tan difundido fue puesto firme e inteligentemente en su lugar por Neissen (1976, págs. 182-185). "El psicólogo no puede predecir la conducta ni controlar a alguien que sabe más... que él acerca de la situación, o que recoge información que el psicólogo no ha tomado en consideración" (pág. 183). De modo que la mejor manera de obtener libertad *práctica* respecto de la predicción y la manipulación de los otros (psicólogos, dictadores o quien fuere) es la adquisición del conocimiento en general. Cuanto más sepamos, más difícil les resultará controlarnos. Lo dice la Biblia: "La verdad os hará libres" —o al menos nos protegerá de ese tipo de sujeción—.

Este es un caso frecuentemente debatido, de acuerdo; pero esta vez consideraré los efectos que producen ciertas palabras muy favorecidas por la tradición. ¿Por qué un neurólogo en lugar de un orador, un maestro o un filósofo? ¿Qué diferencia comporta el hecho de que el neurólogo "manipule directamente" el cerebro de la víctima? No deberíamos descuidar nunca el desplazamiento que sufre el control tanto en los experimentos mentales como en los reales. En los experimentos mentales que son, a la vez, generadores de intuición, esto se efectúa "girando la perilla" —es decir, probando distintas escalas de valores y observando lo que sucede (Hofstadter 1981)—. Vamos a corregir el ejemplo de Fischer:

> Consideremos el caso *rara vez* debatido de un elocuente filósofo que manipula indirectamente el cerebro de una persona, bombardeando sus oídos con palabras de deslumbrante claridad, presentándole razones convincentes y, en consecuencia, induciendo todos sus deseos, creencias y decisiones.

Por cierto, el agente es aun demasiado dominante y no nos gusta, Como señala Nozick (1981, introducción), los filósofos tienen, en última instancia, la desagradable propensión a utilizar armamentos "coercitivos"; aun así, podemos suavizar el cuadro:

> Consideremos el delicioso caso del veraz y bien informado oráculo que manipula indirectamente el cerebro de una persona, bombardeando sus oídos con lúcidos y precisos consejos que resultan irresistibles pues el oráculo no omite citar todas las pruebas que las fundamentan, además del relato sincero del método completo de recolección de pruebas.

Cuando, finalmente, logramos sintonizar nuestro generador de intuición con la imagen de un consejero sabio y amable que sólo utiliza interacciones comunicativas *garantizadas epistémicamente* para producir únicamente efectos cognitivos ("la verdad y nada más que la verdad"), ¿nos seguirá molestando el hecho de que nuestros deseos, creencias y decisiones sean igualmente "inducidos"? Después de todo, es sumamente tranquilizador saber (o tener buenas razones para creer) que es esto y sólo esto lo que nos está sucediendo. El agente nos ha influido, por cierto, pero no nos ha influido "inconscientemente" (Greenspan, 1978, pág. 237). Por el contrario, su influencia depende de, se apoya en y utiliza nuestra inteligencia y nada más que nuestra inteligencia. Nada de "sucias tretas". Tales influencias inducen a decidir *in equilibrium*. Un acto *in equilibrium* resiste el conocimiento de sus propias causas (Nozick 1981, pág. 348). Y lo resiste porque el conocimiento revela que el acto fue

motivado exactamente como lo deseábamos: por una clara visión de la realidad y con la mejor de las intenciones.[23]

Claro que nos gustaría eliminar el intermediario. Preferimos una relación de "primera mano" con el mundo al más perfecto conocimiento "de oídas". ¿Por qué? Porque nos brinda la oportunidad de un aprendizaje que no está condicionado por los intereses epistémicos y los prejuicios de otro agente, y porque disminuyen las probabilidades de engaño: no nos gusta confiar en otros agentes. Nuestra cautela es razonable. Los ogros, si existen, son objetos dignos de temor.

Supongamos que me encuentro en un supermercado tratando de decidir qué clase de sopa voy a comprar. Han llegado a mís oídos rumores muy inquietantes acerca de los trucos utilizados por los publicitarios para "controlar" mi hábito de compra. Ciertamente, la idea me desagrada. ¿Qué espero, entonces? Que al llegar al estante donde están las latas de sopa *nada* controle mi decisión. ¿Quiero la clase de "libertad radical" que me volvería impermeable a los elementos que hacen a la buena calidad del producto? No quiero que mi decisión esté controlada por el color de la lata, por su ubicación en el estante o por el retrato de alguien admirable que me sonríe pegado al envase. Pero si mi decisión estuviese controlada por la calidad, el precio y la disponibilidad del producto, ¿qué mejor influencia podría desear?

Quizá tenga mis razones para exigir un mínimo de "libertad radical", si ese mínimo me protege de los publicitarios embaucadores que pululan por el mundo tratando de "leer mi mente". Así como el piloto evitó una situación en la que corría el *riesgo* de perder el control, gracias a un plan de control de metanivel, del mismo modo nuestro pensar sobre el control de metanivel nos impide utilizar tácticas o estrategias de control demasiado evidentes y que, en consecuencia, pueden ser previstas por un competidor (si alguna vez aparece uno).[24]

[23] Ayer (1954) saca a relucir, y descalifica rotundamente, la curiosa —por no decir ridícula— opinión que define la libertad "como la conciencia de la necesidad". ¿Quizá la idea que se oculta detrás de esta oscura doctrina sea la siguiente: por las razones que ya hemos analizado, tenemos las influencias ocultas que pueden actuar sobre nuestra voluntad. ¿Cuál es el estado que corresponde a la ausencia total de esta terrible condición? El perfecto conocimiento de nuestros deseos y creencias. Por lo tanto, si la libertad es aquello a lo que debemos aspirar, entonces la libertad es la perfecta conciencia de la causalidad (de la necesidad, en un sentido) de nuestras creencias y deseos. La libertad consistirá, pues, en mantener los propios actos en equilibrio y en ser conscientes de ello (en el sentido de Nozick).

[24] Esta razón de control de metanivel que favorece la táctica de la "cara de póker", o sea, la de una conducta impredecible, no necesita ser *nuestro* consciente. Este puede ser una razón flotante, privilegiada por nuestro diseño biológico. Sobre las condiciones bajo las cuales la "cara de póker" es una estrategia estable desde el punto de vista de la evolución... véase Dawkins 1976, págs. 82-83.

5. Los usos del desorden

Nos ocuparemos ahora de un juego muy simple en el que sólo participan dos personas: "piedra, papel y tijeras". En este juego se suele experimentar la sensación de estar encerrado en la cabeza del rival. Uno se familiariza con la pauta de juego del contrincante y gana. Así de sencillo. Imaginemos, pues, que usted está jugando con su amiga María y que le ha ganado veinte jugadas consecutivas de "piedra, papel y tijeras". Acto seguido, María lo invita a abrir el sobre sellado que casualmente dejó sobre la mesa antes de comenzar el juego. Dentro del sobre hay un papel con el siguiente mensaje: "Querido amigo: Trataré de perder veinte jugadas consecutivas de 'Piedras, papel y tijeras', [firmado] María". ¿Quién estaba controlando a quién?[25] ¿Quién tenía el poder? Resulta desagradable enterarse de algo así aunque sea una amiga. Si María puede hacerle una cosa semejante, también podrá hacerlo cualquiera que no sea su amigo.[26] De modo que usted tiene razones para desear que no lean su mente, y si quiere evitarlo, quizá sea necesario no dejar pautas en sus actividades.[27] La única manera de evitar que las pautas no sean legibles es que se vuelvan aleatorias.

Para estar seguros de que ningún contrincante se anticipará a nuestros movimientos, éstos tendrán que depender directamente de alguna indeterminación mecánico-cuántica genuina, es decir, de la clase de libertad que consiste en la ausencia de causalidad. Puesto que, *en principio*, estos efectos cuánticos son impredecibles, también lo serán para la inteligencia superior de Laplace. Por cierto, este tipo de azar no existe necesariamente en el mundo.

[25] La respuesta es obvia si nos referimos al control skinneriano: sus movimientos controlan a los de María. Este caso pone en evidencia lo que hay en la noción común de control que falta en el concepto de Skinner.

[26] "Undercut, Flaunt, Hruska, Behavorial Evolution and Other Games of Strategy" (Hofstadter, 1982c) es un análisis esclarecedor sobre las ventajas y los peligros de adoptar políticas detectables en tales circunstancias.

[27] El errático vuelo de los insectos es un ejemplo de esta estrategia en una razón flotante; en esas condiciones se les hacía muy difícil a los pájaros insectívoros planear una trayectoria que "los condujera al blanco". Pero los pájaros encontraron una manera de compensar el obstáculo aumentando considerablemente el número de parpadeos; de ese modo, verificaban con más frecuencia y precisión la trayectoria de los insectos. Si parpadeáramos con tanta frecuencia como las golondrinas, veríamos una película como si se tratase de una muestra bastante rápida de diapositivas; en otras palabras, el pájaro percibe su mundo "en cámara lenta". Véase Winsatt, 1980. La necesidad que llevó a los pájaros a resolver este problema sintáctico fue la misma que dio origen a la invención de la cibernética. Cuenta Wiener (1948) cómo la solución al problema de dirigir el fuego antiaéreo utilizando medidas ambiguas condujo a la teoría del control y a la cibernética.

Hay innumerables procesos macroscópicos (que son indeterministas a nivel cuántico) que producen resultados aleatorios en un sentido matemático que no tiene nada que ver con la causalidad. En matemática o en teoría de la información, se dice que una serie es aleatoria cuando, tomada como información, no se la puede comprimir, es decir, cuando el camino más corto y más eficiente para "enviar" la serie de A a B es enviarla *verbatim*. Ninguna fórmula o algoritmo para la serie pueden ser especificados en menos *bits* de información que los que lleva transmitir directamente la serie (Chaitin, 1976). (Véanse Berry, 1983 y Ford, 1983.)

Existen, además, series seudoaleatorias similares a las producidas por los "generadores del número aleatorio" en la mayoría de los ordenadores. Estas series no sólo no son causalmente indeterministas, sino que tampoco son incomprensibles desde el punto de vista de la información. Son generadas por algoritmos o funciones simples y breves, pero, "desde afuera" es prácticamente imposible distinguirlas de las auténticas series aleatorias. Esta falta de pautas (o, mejor, esta inescrutabilidad de las pautas) está relacionada con el libre albedrío y con la naturaleza, como veremos en el capítulo 6. Entretanto, nos referiremos a algunas de las funciones que desempeña en nuestras técnicas de control y de autocontrol.

Hemos visto el valor estratégico que tiene el hecho de no dejar pautas en algunas situaciones relacionadas con el juego: nos garantiza que el contrincante no identificará la pauta y, por lo tanto, no la explotará con el fin de ejercer el control sobre nosotros. Además de los juegos contra oponentes reales, existen los "juegos contra la naturaleza".[28] Supongamos que deseamos información acerca de un extenso dominio, pero no tenemos tiempo, ni dinero, ni ganas de verificar cada uno de los elementos del dominio. En ese caso, establecemos un método de "comprobación al azar" que consiste en una pequeña muestra —digamos, el cinco por ciento— cuyos resultados se extrapolan al resto. ¿Cómo elegimos la muestra? ¿Deberíamos escoger el cinco por ciento más "próximo" o el "primero"? ¿Debemos dividir el dominio en un retículo de cuadrados iguales numerados en orden y luego extraer una muestra cada veinte cuadrados: 1, 21, 41, 61, etc.? No. El buen criterio de los estadísticos dice que la muestra del cinco por ciento debe ser extraída al azar. De ese modo, se minimiza el riesgo de utilizar una pauta

[28] Véase también el trabajo de Hintikka sobre la semántica en términos de la teoría de juegos y el abordaje a la inducción por medio de la teoría de los juegos, en Hintikka (1976) y Saarinen (1979).

de muestreo que coincida con la misma pauta que se está buscando en el mundo.[29]

Supongamos que los marcianos envían una sonda espacial a la Tierra para estudiar nuestro clima. La sonda aterriza en el norte de Maine, en el mes de enero y sólo puede extraer 365 muestras antes de que se agoten sus baterías. Si extrajera muestras cada diez minutos no llegaría a febrero (y ni hablar de agosto). Si extrae muestras una vez al día, aproximadamente a la misma hora (a medianoche, como suele suceder), las muestras cubrirán uniformemente el dominio anual, pero no registrarán la existencia de luz diurna. Existen miles de pautas que pueden ser relevantes para la recolección de datos; pero si cualquiera de ellas se encuentra en fase —es decir, si su *ausencia* está en fase— con la pauta de muestreo, resultará invisible. La mejor manera de observar el mundo es desordenar la propia investigación. Una investigación ordenada, sistemática, pautada, será necesariamente ciega para todas las pautas de los fenómenos que se encuentran en fase con la investigación.[30]

El buen criterio implícito en la simple práctica de barajar los naipes antes de comenzar el juego tiene muchas aplicaciones en el mundo de los sistemas diseñados, ya sean biológicos o artificiales. Este buen criterio aparece también en la técnica conocida de invocar la "suerte" arrojando una moneda. Tiramos la moneda para dirimir cuestiones que podrían dirimirse mediante un proceso de deliberación más racional, si bien laborioso. En lugar de decidir racionalmente cuál de los dos equipos *debe* comenzar el partido o a cuál de los dos excelentes restaurantes concurriremos (y de proponérnoslo, encontraríamos razones muy válidas para fundamentar nuestra elección), vamos por la vida permitiendo que una moneda "decida" arbitrariamente por nosotros.

[29] Cuando este libro estaba en prensa, Peddy Seynfeld me hizo reparar en la controversia entre estadísticos y teóricos de la probabilidad sobre el presunto fundamento para la confianza en los procesos aleatorios en los experimentos (un tema lúcidamente analizado en Seynfeld, 1981). Pienso que sus argumentos en contra de las usuales defensas epistemológicas de la práctica son devastadoras; pero las reflexiones que ofrezco aquí y que enfatizan el *desideratum* práctico de minimizar el riesgo de los períodos en fase, son independientes de los argumentos de Seynfeld. No he tenido tiempo de conseguir las pruebas que fundamenten esta importante cuestión, de modo que en esta zona mi bloque de mármol quedará toscamente desbastado.

[30] Por cierto, todo esquema del fenómeno que esté en fase con su esquema de investigación aleatorio o pseudoaleatorio *particular* también será invisible; pero precisamente por la extrema irregularidad de su secuencia de investigación, la probabilidad de un "esquema" igualmente irregular del fenómeno estudiado reflejando perfectamente su "esquema", es infinitesimal. Asimismo, *puede* haber distintas manera de sobreimprimir los esquemas regulares, de modo que estén parcialmente en fase.

Sin embargo, se trata de una práctica racional. Ceder un poco de nuestra racionalidad (de orden inferior) para obtener una mayor eficacia y celeridad en la toma de decisiones es un acto racional (de orden superior).[31] El "coste" es el pequeño riesgo de omitir las razones de peso que justificarían una opción y no la otra. El "beneficio" es evitar el riesgo (mayor) de que expire el plazo para una acción importante a causa de deliberaciones interminables. Es importante, entonces, que la moneda, cuyo comportamiento es utilizado para decidir algo, *no se acople* con lo que constituye el objeto de la decisión —y no sólo para evitar la coincidencia de pautas sino, más bien, el derroche de tiempo que implicaría verificar si dicho acoplamiento es "correcto" (véase Dretske 1981, capítulo 5)—. Si usted desea elegir un restaurante, puede recurrir a un par de dados o bien a un sofisticado *gourmetoscopio* actualizado diariamente con toda la información pertinente. La ventaja de los dados respecto del *gourmetoscopio*, además del precio, es que no necesita mantenimiento. Los dados le responderán cuantas veces le pregunte. ¿Y a quién le importa si la respuesta no es "correcta"?)

6. "¡Suéltese!"

> *No al control o la limitación de uno mismo*
> *para alcanzar fines específicos;*
> *sino un dejarse ir despreocupado;*
> *no la cautela sino una ceguera sabia;*
> *no trabajar para adquirir posesiones*
> *que lentamente crecen, silenciosas,*
> *sino, más bien, el continuo derroche*
> *de todos los valores perecederos.*
> —Rilke, "Über Kunst" [Sobre el arte].

Una vez que la exigencia de actuar —y, por consiguiente, de decidir— en un mundo apremiado por el tiempo pasa a formar parte de nuestra visión del pensamiento racional, obtenemos un modelo

[31] Jaynes (1976) afirma que el descubrimiento de arrojar la moneda (u otra práctica semejante), significó una apertura tecnológica capital en los tiempos prehistóricos y la adquisición de una valiosa técnica cognitiva. En aquel entonces, las personas no resolvían sus problemas de manera compleja y con frecuencia se hallaban en dilemas prácticos cuyos resultados no podían controlar. Para librar a esas personas del dilema del asno de Buridán (que murió paralizado por la indecisión al no poder optar entre la comida o la bebida, ya que sus necesidades respecto de la una y la otra eran equivalentes), se necesitaba lo que Jaynes denomina "métodos exofísicos de pensamiento y de toma de decisión" (pág. 245): sortilegios, echar suertes, examinar las entrañas, "extispicia o aruspicia") y otras prácticas aparentemente supersticiosas. Es interesante que aun hoy necesitemos recurrir a tales estratagemas cuando hay que desempatar.

de la mente radicalmente distinto del modelo clásico (y oscuro) imaginado por Descartes, Leibniz y Kant.[32] Por ejemplo, una cierta dosis de arbitrariedad y de insensibilidad respecto de la información constituye una característica esencial de la racionalidad bajo la presión del tiempo. De lo contrario, no rompería el regreso vicioso de decidir para decidir (sobre la base de fundamentos diversos e incompletos)... para terminar la deliberación y actuar. Cuando elegimos un curso de acción determinado, no lo elegimos porque es el mejor, "luego de haber considerado todos los demás". No sería cuerdo tratar de considerarlos a todos (véase el capítulo 4).[33]

En el capítulo 2 señalé que frente a la eterna seducción del ideal de libertad como *obedeciendo a los dictados de la razón* se halla también la subterránea *repulsión* de la misma idea. En efecto, parece como si en la misma idea de libertad no hubiera espacio para la espontaneidad, el capricho, la individualidad, el arte, etc. Es difícil no caer en la tentación de decir que "si permito que mis actos sean dictados solamente por la razón me vería despojado de mi *personalidad*, transformándome en un mero vehículo de la Verdad o del Justo Obrar y no en un autor único e irrepetible en el escenario del mundo".

Es sencillo descubrir el modelo erróneo de racionalidad que se halla en la fundamentación de esta idea, si lo aplicamos a un dominio mucho más simple que el escenario del mundo: el ajedrez. "Si mis jugadas sólo fueran dictadas por la razón, jugaría siempre el mismo juego pues *el mejor juego es único.*" Quizás hay otro, pero ¿quién puede encontrarlo? Existe, efectivamente, un algoritmo que determina la línea óptima de juego en el ajedrez. Como el ajedrez es un juego finito, se puede, en principio, ubicar todas las jugadas válidas en un gigantesco árbol de decisión y, luego, a partir de las últimas ganadoras, emprender el camino de regreso hasta la mejor apertura. Pero, este árbol de decisión, si bien es finito, contiene más nudos de los que podrá investigar el ordenador más veloz trabajando durante billones de años. La línea más racional que puede seguir una inteligencia finita y macroscópica (que se sabe finita) comportará procedimientos "heurísticos" de decisión que pondrán fin *de manera arbitraria* al modo limitado y arriesgado de análisis, con el objeto de preservar el tiempo de reloj del juego.

[32] Como, por ejemplo, fue propuesto por Simon (1957) en su concepto de "satisfactorio". Para debates filosóficos más amplios sobre la relevancia del concepto de Simon, véanse Nozick (1981) y Cherniak (1983).

[33] Diseñar un método criterioso y funcional para ignorar las cosas ha resultado ser uno de los problemas más profundos y difíciles en inteligencia artificial. Véase Dennett, en preparación-b.

Teniendo en cuenta estas limitaciones, la línea más racional a seguir es no jugar siempre el mismo juego. Esto se funda en varias razones. Primero, usar siempre el mismo plan de juego genera un patrón de juego que el contrincante podrá aprovechar: en sus horas libres, analizará las jugadas hasta romper ese modo particular de ataque, seguro de que sus esfuerzos no serán vanos. Segundo, "barajar" las decisiones en medio del juego produce, en cierto modo, una mayor "variación genética", a partir de la cual se pueden elegir y descubrir estrategias nuevas y más eficaces. Estas estrategias libres pueden ser privilegiadas simplemente porque el agente fue *diseñado* para privilegiarlas, valorarlas y disfrutar de ellas. Seguramente, existe una razón poderosa (flotante o no) en favor del diseño para un agente capaz de apreciar la variedad de aburrirse de las repeticiones (pero, también, de dejarse cautivar por los patrones de juego).

En el ajedrez nos encontramos a menudo con situaciones sin salida, semejantes a las que observamos en el más vasto y problemático juego de la vida. Existen, por ejemplo, las "jugadas forzadas". En ocasiones, las reglas del ajedrez fuerzan las jugadas: en tales instancias, uno se encuentra ante el hecho irrevocable de que sólo es posible una sola jugada legal (si uno no está en jaque y no es posible ninguna jugada legal, hace tablas). Pero nos interesan mucho más las jugadas forzadas en las que se puede optar entre dos o más jugadas legales, pero donde sólo una de ellas no es tonta ni "suicida". Por esta razón, decimos que se trata de una jugada forzada, no por las reglas del ajedrez ni por las leyes de la física, sino por los dictados de la razón. Es, obviamente, lo único que podemos hacer, dado nuestro interés en ganar (o al menos en no perder) la partida.

Solemos desesperarnos cuando nos hallamos en una situación en la cual sólo hay una única solución racional. A veces, es otro agente quien nos pone en esa situación: "Le hice una oferta que no pudo rechazar". Otras veces es la Madre Naturaleza la que, a su manera impersonal, nos acorrala. En este caso, la supervivencia depende de nuestra habilidad para seguir una trayectoria particular (y previsible). Por ejemplo: si usted no encuentra un refugio cálido dentro de una hora, morirá por congelamiento. El único refugio accesible se halla del otro lado del precipicio y la única forma de atravesarlo es pasar por un puente colgante hecho con sogas, etc. Nos asisten razones para temer tales situaciones límite, en las cuales nuestras opciones se reducen a una sola. Preferimos contar con muchas opciones, hasta el punto de empantanarnos por el exceso y la falta de tiempo para examinarlas a todas. Preferimos las circunstancias en las que podemos permitirnos un mínimo de exploración espontánea y, aunque se trate de una preferencia quizás

estimulada genéticamente o profundamente arraigada en nuestro carácter, también está dictada por la razón.

El conflicto enunciado por Rilke entre el ideal de espontaneidad y el ideal de deliberación racional es ilusorio. Ninguna persona elegirá una política de deliberación y decisión que sea completamente caótica. Asimismo, el más puro devoto de la razón no escogerá, a causa de esa misma devoción, una política que excluya una buena cuota de espontaneidad. Como veremos en los últimos capítulos, dicha política también se manifiesta en el modelo de nosotros mismos como agentes libres, en la "idea de libertad" bajo la cual actuamos y en la política de *no mirar demasiado de cerca* cuando nos observamos u observamos a los otros como agentes libres.

Habría mucho que agregar acerca del autocontrol y del concepto del yo, pero hemos ido bastante lejos en nuestro examen y es tiempo de exponer las conclusiones tranquilizadoras. Contrariamente a la imagen familiar que presenté al principio del capítulo, *el determinismo en sí mismo no "erosiona el control"*. El cohete espacial Viking es tan determinista como puede serlo un reloj, sólo que se controla a sí mismo. Dispositivos deterministas más complejos no solamente se controlan a sí mismos, sino que evitan cualquier intento de ser controlados. Si somos dispositivos deterministas, nos controlamos y controlamos nuestros destinos; por lo tanto, no hay nada que temer al respecto.

Más aun, *el pasado no nos controla*, así como la NASA no controla las naves que vagan por el espacio, fuera de su alcance. Ello no significa que no existan lazos causales entre la Tierra y esas naves, pues los hay. Pero esos lazos causales no bastan para controlarlos: debe de haber también retroalimentación para poder informar a quien controla. Y no hay señales de retroalimentación del presente hacia el pasado para que este último las utilice. Además, no hay nada en el pasado que prevea y planee nuestros actos particulares, aun cuando fuera cierto que la Madre Naturaleza, anticipando nuestras necesidades y atolladeros, nos diseñó para cuidarnos muy bien. No estamos controlados por nuestros ancestros ni por nuestro pasado evolutivo, sino que, por el contrario, la herencia ha tendido a constituirnos como seres que *se controlan a sí mismos* (afortunadamente).

Lo que comienza a perfilarse en esta investigación es el esbozo de un ideal: cómo nos gustaría ser. Nos gustaría ser inmunes a la manipulación y a los trucos sucios, y sensibles a los presagios de vicisitudes futuras, a fin de poder alterar correctamente el curso de los acontecimientos y, así, enfrentar al mundo con tanta libertad de acción, con un margen tan amplio de error y una cuota tan pequeña de incertidumbre como seamos capaces de conseguir.

Esto implica que desearíamos que el mundo fuera también de cierta manera: variado, pródigo en sustento y placer y —lo más importante en este contexto— no tan exigente (dadas nuestras habilidades y necesidades) como para restringir brutalmente nuestras opciones al mínimo. Si nos toca vivir en una isla desierta, la haremos fértil y rica en posibilidades y, de ninguna manera, tan árida e inconmovible que nuestras jugadas sean forzadas como en el ajedrez.

La tesis sobre el determinismo no tiene, por cierto, implicaciones respecto del modo como el mundo se ordenará a sí mismo, sin embargo, solemos responder a esta perspectiva como si se tratara de un futuro que habrá de limitarnos. Como si el piloto del avión nos dijera por la radio: "Me temo que de ahora en adelante entraremos en una vasta zona de *determinismo*. No tendremos *poder* ni *libertad de maniobra*. ¡Afortunados de nosotros, si sobrevivimos intactos!" Esta no es, ciertamente, la forma correcta de enfocar el problema, pero aún no hemos indagado a fondo ni descalificado todas las razones que nos inducen a pensar que hay un grano de verdad en todo ello.

4

El yo que se hace a sí mismo

Por mi parte, cuando penetro más
profundamente
en lo que llamo mi yo,
siempre tropiezo con una
u otra percepción particular
de calor o de frío, de luz o de sombra,
de amor o de odio, de dolor o de placer.
Nunca puedo atraparme a mí mismo
sin que una percepción me acompañe en todo
momento
y nunca puedo observar sino esa percepción.
—Hume, *Tratado de la naturaleza humana*, I, IV, iv.

Hay ciertas almas que nunca se descubren,
a menos que uno las haya inventado primero.
—Nietzsche, *Así hablaba Zaratustra*, parte II.

1. El problema del yo que desaparece

La primera vez que estuve en Londres, mientras buscaba la estación del metro más próxima, di con una escalera que llevaba el cartel de *Subway*, un término que en Boston equivale a *underground* ["bajo tierra"]; de modo que bajé muy confiado y seguí caminando en busca de los trenes. Luego de errar por galerías y pasillos, encontré otra escalera, subí con ciertas dudas y aparecí en la acera opuesta al sitio de donde había partido. "Debo de haberme salteado una vuelta", pensé, y me lancé escaleras abajo para intentar otra vez la cosa. Después de lo que me pareció una búsqueda exhaustiva de

molinetes y entradas laterales hasta el momento inadvertidos, aparecí nuevamente en la acera del principio, sintiéndome la víctima de una broma pesada. Finalmente, caí en la cuenta: estaba cometiendo un error de categorías. Algo similar ocurre cuando uno busca el yo o el alma. Uno entra en el cerebro por el ojo, camina por el nervio óptico, da vueltas por la corteza, mirando detrás de cada neurona y, cuando uno menos se lo espera, sale a la luz del día en la punta del impulso de un nervio motor, rascándose la cabeza y preguntándose dónde está el yo.[1]

Hasta el momento, en nuestra discusión sobre el autocontrol, no hemos dicho qué es el yo, dónde se encuentra y de qué se compone (si acaso está compuesto de algo). ¿Hay alguna razón que explique la presencia del yo entre las cosas existentes?[2] El conocido fracaso de Hume al intentar encontrar el yo mediante la introspección es análogo a otras incursiones que terminaron con las manos vacías: la búsqueda de Mozart de la fuente primordial de su genio creador (véase el capítulo primero, pág. 26) y la búsqueda del "delincuente interior" cuando se trata de atribuir culpas. Nagel señala:

> Si uno no puede ser responsable de las consecuencias de sus actos debido a factores que están fuera de su control, o porque los antecedentes de sus actos son propiedades del temperamento que no están sujetas a la voluntad, o debido a las circunstancias que suscitan las propias elecciones morales, entonces ¿cómo se puede ser responsable incluso de los actos desnudos de la voluntad misma? El área de la auténtica agencia y, en consecuencia, del legítimo juicio moral parece encogerse, cuando se los examina, hasta un punto sin extensión (Nagel 1979, pág. 35).

Este punto sin extensión no es, desde el punto de vista teórico, muy prometedor como modelo del agente moral. En este pasaje, Nagel parte de algunos supuestos acerca de las condiciones en las cuales es posible controlar ciertas cosas. (En el capítulo 3 mostramos que dichas condiciones eran falsas.) Pero, como el control es sólo un aspecto de la responsabilidad, incluso una explicación causal del control más cauteloso puede dejar al *yo* en peligro de disolución

[1] Algunos escritores piensan que existen puertas secretas en el laberinto del cerebro por las cuales pasan los mesajeros del yo, con la ayuda de una llave muy especial y muy misteriosa. Véase, por ejemplo, Popper y Eccles 1977. Esa no es la manera de localizar el yo.

[2] *The Mind's I* [El ojo de la mente] (Hofstadter y Dennett 1981) está dedicado a la busca del yo; de ahí mi breve discusión sobre este importante tema. Respecto de la pregunta acerca de la existencia del yo, véase mi introducción (págs. 3-16) y las "Reflexiones" sobre Borges (págs. 348-352). Véase también Dennet, en preparación-b.

"a través de la erosión gradual de lo que hacemos por la sustracción de lo que sucede", como dice Nagel (1979, pág. 38). *¿Hacemos* algo? ¿Qué significa ser un agente y no una mera hebra en la tela de la causalidad?

> Pero usted se había embarcado en una ilusión más grande y más dañina. La ilusión del control. De que A pudiera hacer B. Pero es falso. Absolutamente. Nadie puede *hacer*. Las cosas simplemente suceden... (Dicho por un personaje de la novela de Thomas Pynchon, *Gravity's Rainbow*, 1973, pág. 34.)

Por una parte, tenemos la imagen del agente, del hacedor; el *locus* y la fuente de la acción y no de la mera reacción. Rawls señala que "los ciudadanos se conciben a sí mismos como *fuentes autocreadoras* de reclamos válidos [la bastardilla es mía]". Todos querríamos decir de nosotros mismos lo que dijo Harry Truman: "Yo asumo toda la responsabilidad". Por otra parte, tenemos la imagen física del ser humano, que lo presenta como algo tan insignificante como la ficha de dominó número 743 de una serie compuesta por millones de fichas. Si somos meros conductos de causalidad, no podemos ser, al mismo tiempo, agentes.

Esto conduce de manera inevitable a la doctrina de la "Causalidad del agente". Reid dice: "En el presente caso, o bien el hombre era la causa de la acción y entonces se trataba de una acción libre que le fue justamente imputada; o bien hubo otra causa y no puede ser imputada al hombre sin cometer injusticia" (Reid 1788, pág. 329). La teoría de la causalidad del agente es misteriosa, aunque tradicional. Chisholm, su principal exponente, la expresó en los siguientes términos. (Véase también Chisholm 1976, especialmente el capítulo siete.)

> Si somos responsables... entonces poseemos una prerrogativa que muchos atribuirían solamente a Dios: cuando actuamos, cada uno de nosotros es un primer motor inmóvil. Al hacer lo que hacemos, causamos ciertos acontecimientos, y nada —o nadie— es la causa de que causemos esos acontecimientos (Chisholm 1964b, reeditado en Watson 1982, pág. 32).

¿Podemos encontrar una descripción naturalista del yo o del agente que evite esta "metafísica oscura y aterradora" [3] y que, al mismo tiempo, distinga suficientemente a los agentes dentro del entramado de la causalidad? Esta es la tarea del presente capítulo.

[3] Strawson (1962) se refería a la metafísica del "libertarismo", pero por esta vez yo querría extender el epíteto a la metafísica de la Causalidad del Agente.

Es evidente que tal descripción positiva tendrá que declarar que las intuiciones que sustentan la visión chisholmiana del yo como motor inmóvil son una mera ilusión cognitiva. Dicha declaración deberá fundarse en un diagnóstico convincente de la causa de esta ilusión. Hasta cierto punto, *parece* como si fuéramos motores inmóviles; pero si se trata de una *mera* apariencia, ¿qué hace que esto parezca ser así? No uno, diría, sino varios factores.

En primer término, el sistema nervioso magnifica los efectos produciendo algo semejante a una ilusión en las proporciones. Entre otras cosas, somos sistemas que procesan información y todo sistema que procesa información depende de distintos tipos de amplificadores. Causas relativamente pequeñas producen efectos relativamente grandes. En particular, el gasto de grandes cantidades de energía almacenada se controla por la modulación de los acontecimientos de más baja energía. Un sistema de procesamiento de información es esencialmente una organización de *interruptores* o *disparadores*. Los interruptores que controlan el *output* de los ejecutores —brazos, piernas, bocas, ruedas, proyectiles, misiles— utilizan muy poca energía de *input* para iniciar, modular y terminar los procesos o actividades que insumen dramáticas cantidades de energía, como cualquiera puede observar.

Por otra parte, los interruptores de *input* del sistema —los transductores que forman los órganos de la percepción— también son amplificadores. El "encendido" de una neurona retiniana, por ejemplo, puede ser "disparado" por la llegada de un único fotón al receptor de la retina. Se aprovecha una gran cantidad de información con cantidades insignificantes de energía y, entonces, gracias a los poderes de amplificación de los sistemas de los interruptores, la información comienza a realizar su trabajo —evocando otra información que fue almacenada hace mucho tiempo y transformándola, para la ocasión, en un millón de pequeños senderos que eventualmente conducen a una acción cuya genealogía de causas eficientes es tan desesperadamente inescrutable que se *vuelve* invisible—.[4] Vemos los efectos que salen, no vemos las

[4] En semejante enredo de microcausalidad uno podría postular, sin temor a incurrir en contradicciones experimentales, la posible contribución de los acontecimientos indeterministas del nivel cuántico. La mera ubicuidad de los efectos cuánticos microscópicos en el mundo y, por consiguiente, en nuestros cerebros, no nos aseguraría un indeterminismo macroscópico de la acción humana, a menos que hubiera transductores internos que fueran, en efecto, contadores Geiger naturales en miniatura, capaces de amplificar el efecto del electrón a la escala en que ocurre la interrupción del control del sistema nervioso. De lo contrario, las perturbaciones aleatorias en el nivel subatómico serían absorbidas como ruido de fondo, no teniendo ningún efecto sobre la

causas que entran; nos tienta la hipótesis de que existen causas.[5]

Esta invisibilidad de los senderos causales no es similar a la invisibilidad (para nosotros) de las otras mentes. Desde la perspectiva "introspectiva" de la primera persona, los senderos causales son igualmente inhallables.[6] Consideremos las decisiones en sí mismas, tal vez los acontecimientos centrales en la vida mental de un auténtico agente libre. ¿Se trata de decisiones voluntarias o de cosas que meramente nos suceden?[7] Desde cierto punto de vista, parecen ser los movimientos esencialmente voluntarios de nuestra vida, los instantes en que ejercemos en plenitud nuestra facultad de actuar. Sin embargo, se puede considerar que esas mismas decisiones están fuera de nuestro control. Tenemos la esperanza de observar cómo decidimos algo, y cuando realmente lo decidimos, nuestra decisión ya ha aflorado a nuestra conciencia,

---

disposición de los interruptores del sistema de procesamiento de información. Describir estos mecanismos es fácil; la dificultad estriba en encontrarles un objetivo (a la luz de las discusiones de los usos de la mera pseudoaleatoriedad en el capítulo 3).

[5] Aristóteles en la *Etica a Nicómaco* (Libro III) caracteriza al acto voluntario como aquel cuya fuente está dentro del agente, y al acto involuntario como "aquel cuyo principio de movimiento se encuentra fuera de él". En la *Física* (Libro VII) afirma que todo lo que está en movimiento es movido por sí mismo o bien por otro, señalando que un agente capaz de actos voluntarios debe pertenecer a la clase de los semovientes. Pero Thorp (1980) muestra que el análisis de Aristóteles sobre el concepto de semoviente lleva eventualmente a la conclusión de que cualquier semoviente no es realmente simple sino que está compuesto por partes, algunas de las cuales mueven a otras partes y obtienen movilidad desde afuera. Para Thorp, esto es "decepcionante", pues muestra "la absoluta insuficiencia, para los libertarios, de la explicación del automovimiento de Aristóteles" (pág. 99). Según Aristóteles: "el automovimiento no es una categoría estricta sino una especie de engaño" (pág. 98); uno de sus defectos, señala Thorp, es que un avión a control remoto podría ser considerado semoviente (pág.99).

[6] Hume es particularmente sagaz respecto de la engañosa fenomenología de la decisión: "Sentimos que nuestras acciones están sujetas a nuestra voluntad en la mayoría de las ocasiones, e imaginamos sentir que la voluntad misma está sujeta a nada; porque cuando al negarla, somos llevados a intentarlo, sentimos que se mueve fácilmente por todas partes, y produce una imagen de sí misma incluso en ese lado donde no se estableció" (*Tratado II*, Selby-Bigge, pág. 408).

[7] Ryle, en "The Mith of Volitions" (Ryle 1949, págs. 62-68) ofrece una *reductio ad absurdum*: si las acciones voluntarias son aquellas producidas por las *voliciones* de un agente (sea quien fuere) ¿se trata entonces de voliciones en sí mismas voluntarias? Si es así, obtenemos un infinito regreso; si no es así, obtenemos actos voluntarios como resultado de acontecimientos involuntarios (o no voluntarios) —"una conclusión que podría molestar a esos moralistas que usan la volición como la tabla de salvación de sus sistemas"—(pág. 67). Uno puede coincidir con Ryle en burlarse de las voliciones en tanto entidades teóricamente poco agraciadas para la psicología y además suponer que las decisiones son ítems suficientemente familiares como para que los teóricos confíen en ellas; yo estoy planteando problemas acerca de la categoría ordinaria de las decisiones.

venida no sabemos de dónde. No hemos presenciado su *formación*; hemos presenciado su *llegada*. Esto nos conduce a la extraña idea de que el cuartel general de operaciones no está donde nosotros (como conscientes introspectores) estamos, sino en alguna parte más profunda y a la cual no tenemos acceso. La famosa pregunta de E.M. Forster "¿cómo puedo descubrir lo que pienso hasta que no vea lo que digo?" parecen las palabras de un extranjero esperando un comunicado del interior.

¿Pero por qué interpretar nuestro imperfecto autoconocimiento (nuestro "acceso subprivilegiado", en palabras de Gunderson) de esta manera, como una literal *excentricidad* de nuestra posición de ventaja con respecto al propio autocontrol? ¿Por qué debe existir un centro? La ilusión de un centro último surge cuando tomamos una idea buena, la idea del yo, como un punto de vista coherente y unitario sobre el mundo (Dennett, en preparación-b), y la empujamos demasiado lejos bajo la presión de las preocupaciones acerca de nuestra responsabilidad (no sólo moral, sino también intelectual y artística, como en el caso de Mozart). Nos preguntamos: "¿Realmente *hice* eso?", y nuestro intento por contestar la pregunta provoca algo parecido a una construcción geométrica en busca de una interpretación.

De acuerdo con el concepto tradicional de responsabilidad, somos primaria y directamente responsables de nuestras acciones "voluntarias", esto es, de las cosas que "hacemos" y (en la mayoría de los casos) sólo indirectamente responsables de las cosas que nos suceden. Se sostiene, por ejemplo, que "no puedo evitar" la ira (o la lujuria) que experimento cuando los objetos del entorno se presentan ante mis sentidos de cierta manera. Sin embargo, se supone que gobierno, con mi fuerza de voluntad, los pensamientos y las acciones consiguientes referidos a esos objetos. Cuando nos observamos introspectivamente con el propósito de agrupar nuestros acontecimientos mentales en estas dos categorías morales, nos sucede algo peculiar. Los acontecimientos próximos a las "periferias" de *input* y *output* se adecuan sin problemas a dichas categorías. *Sentir dolor en mi pie* o *ver el escritorio* no son actos que están "bajo mi control" sino cosas que me suceden como resultado de los estímulos o impactos del mundo; y *mover mi dedo* o *decir estas palabras* son obviamente cosas que hago —acciones voluntarias por excelencia—.

Pero, cuando nos alejamos de la periferia hacia el presunto centro, los acontecimientos que tratamos de examinar se nos aparecen extrañamente fluctuantes. La pasividad de la percepción ya no nos parece tan obvia. ¿No contribuyo en algo a mi percepción, voluntariamente, y aun a mi reconocimiento o "aceptación" del

escritorio en tanto escritorio? Después de todo, ¿puedo *no suspender el juicio* ante lo percibido e impedir la certeza? Y del otro lado del centro, cuando observamos más de cerca la acción, ¿mi acto voluntario está realmente *moviendo* mi dedo o más bien *tratando* de mover mi dedo? El conocido generador de intuiciones acerca de las acciones voluntarias de alguien totalmente paralizado prueba que no poseo el control (ni en el mundo ni en mi cuerpo) de todas las condiciones necesarias para que mi dedo realmente se mueva.

Ahondando un poco más, ¿es tan obvio que *aceptar* o *tratar* son acciones voluntarias? Quizá no sea ilusorio que la aceptación de una proposición pueda ser concedida o denegada "a voluntad".[8] Hay condiciones en las que soy incapaz de hacer tal o cual cosa porque las mismas condiciones están fuera de mi control.

Ante la incapacidad de "ver" introspectivamente dónde está el centro o la *fuente* de nuestras acciones libres y ante el rechazo a abandonar la certeza de que somos responsables de lo que hacemos, aprovechamos el vacío cognitivo —es decir, las lagunas de nuestro autoconocimiento— llenándolo con una entidad mágica y misteriosa: el motor inmóvil, el yo activo. Este salto teórico se pone en evidencia en nuestras reacciones ante las fallas de la "fuerza de voluntad". "Me levantaré de la cama y me pondré a trabajar ahora mismo", suelo decirme, y continúo somnoliento en la cama. ¿He tomado o no la decisión de levantarme? Quizá creí haber tomado una decisión. Una vez que reconocemos que el acceso consciente a las propias decisiones es harto problemático, podemos darnos cuenta de que muchos de los momentos decisivos de nuestra vida —hasta donde llega la memoria de la experiencia consciente— no estuvieron acompañados por decisiones *conscientes*. Cuando uno se dice: "He decidido aceptar el trabajo", uno está seguro de que se está informando a sí mismo de algo que ha hecho recientemente; pero la reminiscencia sólo muestra que ayer uno estaba indeciso y que hoy ya no lo está. La decisión *debe haber sucedido* en algún momento del intervalo, sin fanfarrias. ¿Dónde sucedió? En el cuartel general de operaciones, por supuesto.

Pero tal deducción revela que estamos construyendo una teoría psicológica de la "decisión", idealizando y extendiendo nuestra práctica real e insertando decisiones donde las requiere la teoría y no donde tenemos experiencia directa de ellas. Debo haber tomado una decisión, razona uno, puesto que veo que estoy absolutamente

[8] El supuesto voluntarismo de aceptar o rechazar proposiciones juega un extenso y problemático papel en la teoría de la decisión y en las disciplinas normativas relacionadas con ésta. Véase, por ejemplo, Eells (1982, cap. 2).

decidido y ayer no lo estaba. El misterioso santuario interior del agente central comienza a hacerse cargo de la propia vida teórica y a preparar el terreno para la doctrina de los agentes en tanto motores inmóviles.[9]

Tales extravíos difícilmente podrían sostener por sí solos la doctrina de la Causalidad del Agente, pero pueden contribuir de manera espuria al atractivo de la doctrina si van acompañados por la convicción de que no se puede dar ninguna explicación naturalista positiva de aquello que amamos cuando contemplamos "el amado yo". Un modo de socavar esta convicción es aventurar una explicación positiva, y eso es lo que haré acto seguido.

2. *El arte de la autodefinición*

Una cosa es necesaria
para "dar estilo" al propio carácter
¡un arte grandioso y raro!
—Nietzsche, *La gaya ciencia*

En primer término, me ocuparé de algunos aspectos, generalmente soslayados, del proceso por el cual nos convertimos en un yo. La correcta evaluación de estos aspectos nos permitirá librarnos de uno de los temores fundamentales que origina el problema del libre albedrío. En particular, examinaré el papel desempeñado por la *suerte* en el desarrollo del yo; el examen requerirá un breve análisis previo del concepto de suerte (un concepto engañoso, que nos conduce a muchas trampas). Esto brindará una explicación de nuestro acceso a la personalidad moral plenamente madura, pero debe ser protegida de la clásica guerra filosófica que amenaza con

[9] Schneewind (1981) señala que la autobiografía de John Stuart Mill es maravillosamente coherente con su teoría compatibilista-determinista de la acción.

> Mill nos da un ejemplo fascinante de ... un intento determinista de retratarse a uno mismo desde adentro, sin abandonar la teoría filosófica o perder todo el sentido de vivir una vida individual vivida ... La famosa "crisis" en la historia mental de Mill es sencillamente "la experiencia de las penosas consecuencias del carácter que [él] tenía anteriormente"; y el amor por Harriet Taylor es el "fuerte sentimiento de admiración por el anhelo accidentalmente despertado". Mill nos presenta ambos acontecimientos —los dos momentos críticos de su vida— en un lenguaje que evita toda insinuación de una posible elección, decisión o voluntad (pág. 1234).

No quiere decir que un compatibilista deba negar el hecho de que siempre se toman decisiones —y Mill, señala Schneewind, alude a varias decisiones y elecciones "triviales"— sino que un compatibilista como Mill no está bajo ninguna presión teórica que lo lleve a vincular un disparador metafísico particular con cada episodio moralmente relevante en su vida.

desgarrarla: por un lado, el exceso de simplificación, el cual, naturalmente, engendra el escepticismo; por el otro, la exageración compensatoria, la promulgación de doctrinas absolutistas para "proteger" la idea del yo del mecanismo usurpador.

Un yo, ante todo, es un lugar de autocontrol. Cuando se mira al espejo, ¿cómo sabe usted que es la persona que se mira al espejo? Notar el parecido no basta, pues usted no sabe cómo es exactamente en este preciso instante.[10] Las pruebas para determinar si se está mirando a sí mismo en el espejo son las mismas que se utilizaron en el caso del avión de juguete. Usted tratará de comprobar si controla los movimientos de la persona que está viendo. ¿Puede mover *esa* mano? Si puede, es la suya.[11] El problema gnoseológico de este modo de autoidentificación es normalmente trivial pero se complica en casos más artificiales. En un inextricable enredo de brazos y piernas convenientemente disfrazados —digamos, en un grupo de lucha con trajes y guantes idénticos—, la pregunta "¿cuál de estas manos es la mía?" puede requerir cuidadosos experimentos y reflexiones.[12] El criterio último es el control: soy la suma total de las partes que controlo directamente.

¿Todos los yo se controlan a sí mismos? En ciertos aspectos, todos los organismos lo hacen. Su conducta no está (normalmente) controlada por algo o por alguien y está controlada principalmente para beneficio de dicho organismo.[13] Los organismos fueron dise-

[10] La mayoría de los animales aparentemente no saben cómo es su apariencia —¡a menos que se trate de un conocimiento innato!— Pero algunos animales, especialmente ciertos monos, pueden llegar a reconocerse en el espejo. Y cuando lo hacen, el orden de su razonamiento es muy claro: puesto que la imagen es mía, yo debo ser así (debo tener una mancha de tizne en la frente). Véase Gallup, 1977.

[11] La relación entre autocontrol y referencia "indexical" aparece (pero no se discute) en los incontables ejemplos que ilustran la reciente literatura sobre el tema. Véase Böer y Lycan para una discusión útil y una extensa bibliografía de los últimos trabajos aparecidos. Véase también Hofstadter y Dennett 1981, *passim*, y especialmente las páginas 20-22, 265-268, 471.

[12] Hace algunos años, el puerto de New York probó un sistema de radar compartido por los propietarios de pequeños botes. Una única y poderosa antena formaba la imagen de radar del puerto que podía entonces ser transmitida como una señal de televisión a los propietarios de los botes, que, de esa manera, podían ahorrarse los gastos del radar, instalando simplemente un aparato de televisión. ¿En qué consistía la bondad de tal sistema? Si usted estaba perdido en la niebla y miraba la pantalla de televisión, sabía que uno de los muchos puntos en la pantalla era usted, pero ¿cuál de ellos? Realice una simple prueba. Haga girar su bote en un estrecho círculo; entonces su punto en la pantalla será el que traza la letra "O" (a menos que varios botes perdidos en la niebla hagan la misma prueba al mismo tiempo).

[13] Véase Dawkins 1982, capítulos 11-14. Se obtienen sorprendentes resultados al adoptar la perspectiva de observar la conducta de los animales como si estuvieran bajo el control y trabajaran para el beneficio de otros (animales o genes).

ñados para cuidarse a sí mismos. Pero, aun cuando estén dotados de una suerte de yo, nuestra intuición nos dice que se trata de un yo en extremo rudimentario y chato si se lo compara con el nuestro.

Para comprender mejor la diferencia entre nuestro yo y el de las criaturas inferiores, nos serviría de gran ayuda un cuento sobre los aspectos críticos del proceso que va desde las criaturas inferiores y llega hasta nosotros. Una parte de esta historia se encuentra, seguramente, en el cuento del *homo sapiens* (que narra cómo apareció el *homo sapiens*, una vez que el entorno empezó a incluir las complicaciones sociales que habían creado sus ancestros). Pero también es evidente que una parte importante del desarrollo de la persona ocurre en el individuo. Intuitivamente hablando, los lactantes apenas tienen un yo o lo tienen menos desarrollado que los gatos y los perros. Ciertos procesos de aprendizaje y de maduración los convierten en los adultos morales autoconscientes que pretenden ser.

La idea de que un yo plenamente desarrollado o *persona* es el resultado de procesos de interacción social ya ha sido hábilmente investigada por otros filósofos y no perderé el tiempo en revisar sus argumentos.[14] Más bien, enfocaré el proceso de adquisición del yo desde la perspectiva más simple de las estrategias de control.

El individuo humano comienza la vida con un problema: aprender a controlarse a sí mismo. Como en todos los problemas, se parte de una determinada posición, con determinados recursos, que luego deben ser ajustados y combinados para obtener una solución. Esto puede sonar trivial, pero contradice una teoría que, si bien ha gozado de una enorme influencia, es, en última instancia, incoherente: la teoría de la autoelección de Sartre; es decir, la idea de que uno se elige a sí mismo *totalmente* mediante una "elección radical" que no lleva consigo ningún lastre del pasado. Como si uno pudiera crearse a sí mismo *ex-nihilo*; como si, para usar otra expresión latina cara a los filósofos, el yo fuera una *tabula rasa* en el momento de nacer. Sabemos que no es así.

Por cierto, Sartre también lo sabía, pero eligió esta metáfora extravagante para expresar su concepción.[15] No se trata de una metáfora útil. Postula la existencia de un poder mágico que ninguno de nosotros posee y nos induce a creer que, si la ciencia demuestra

[14] Véase Strawson 1959 para un análisis filosófico ingenioso y especialmente abstracto de esta declaración, y véase Jaynes 1976 para una imaginativa reconstrucción de las fuerzas y presiones históricas que pudieron haber dado forma a la evolución sociobiológica. También resulta muy útil Campbell 1975 y A. Rorty 1976.

[15] No sé cómo hacer una justa exégesis de los dos pasajes en una única página de Sartre: "De modo que en cierto sentido *elijo* haber nacido" y "En realidad, soy responsable de todo, excepto de mi responsabilidad misma, pues yo no soy el fundamento de mi ser" (Sartre 1959, pág. 555).

que tal poder no existe, no podemos ser libres. No sólo sabemos empíricamente que un recién nacido no es una pizarra en blanco, sino que también sabemos, por un razonamiento *a priori* tan coercitivo como cualquier otro, que es *así* como debemos comenzar si hemos de ser capaces de aprendizaje o desarrollo. No es difícil adivinar por qué un filósofo como Sartre querría aferrarse al absoluto, ya que existe un argumento que con frecuencia aflora a la superficie: uno no puede ser responsable en ningún sentido, a menos que sea absolutamente responsable de sí mismo.

Los "compatibilistas" o "deterministas blandos" —aquellos que creen que el libre albedrío y la responsabilidad son compatibles con el determinismo— sostienen que uno actúa libre y responsablemente sólo cuando hace lo que ha decidido, basándose en lo que cree y desea. Pero los "deterministas duros" y otros escépticos consideran esto como una mera postergación de las dificultades. Edwards nos brinda una lúcida versión del argumento básico:

> Supongamos que A y B son personas compulsivas que sufren intensamente por sus respectivas neurosis. Supongamos que existe una terapia que puede ayudarlos, cambiando materialmente la estructura de sus caracteres, pero que insume una enorme cantidad de energía y de coraje. Supongamos que A tiene el coraje y la energía necesarios, pero que B no los tiene. A se somete a la terapia y se produce el cambio deseado; B se vuelve cada vez más compulsivo y desdichado. Ahora bien, es cierto que A ayudó a modificar la estructura de su carácter, pero el punto de partida, su deseo de cambiar, su coraje y su energía ya existían. Quizás fueron el fruto de esfuerzos previos de su parte. Pero tuvo que haber un primer esfuerzo y en aquel momento fue el resultado de factores que no dependían de A. (Edwards, 1981, pág. 121.)

Esta objeción es más sutil que nuestra preocupación de que el pasado nos controle. Quizás el pasado no nos controla, pero mi capacidad actual de tomar decisiones y de controlarme a mí mismo no es, aparentemente, algo de lo cual pueda considerarme responsable. ¿Cómo podría un proceso determinista de "transformación del carácter" que parte de un ser que *no* es responsable de ninguna de sus "decisiones" producir un ser que no sólo fuera responsable de sus decisiones, sino también del tipo de carácter que formaría esas decisiones? ¿Cómo podría un proceso *indeterminista* producir algo mejor, a partir del mismo punto de partida?

A menos que encontremos la manera de producir un yo responsable a partir de elecciones inicialmente no responsables, de modo que haya una adquisición gradual de responsabilidades por parte del individuo, nos quedaremos aprisionados en una alternativa

difícil de digerir. O bien negamos el tan mentado yo responsable, o bien adherimos a la solución de Sartre y Chisholm y abrazamos una doctrina francamente misteriosa, que podríamos llamar la doctrina del *agente absoluto*.[16] Para el sentido común la responsabilidad se adquiere —o se gana— gradualmente; pero, a menos que podamos encontrar una falla fatal en el argumento de Edwards, el sentido común está en peligro y nuestra lealtad hacia él no será otra cosa que una expresión de deseo.

Antes de retomar la explicación positiva de la adquisición del yo responsable, podemos darnos un respiro mientras observamos que el argumento de Edwards es sospechosamente similar a un caso de la bien conocida pero mal comprendida falacia llamada *sorites*. (Un ejemplo típico: usted no conseguirá nunca que un hombre que no es calvo se vuelva calvo arrancándole solamente un pelo de la cabeza. Ahora inténtelo, arránquele un pelo; ahora otro; después un tercero y así sucesivamente. Según nuestra primera premisa, el hombre nunca se quedará calvo ¡no importa cuántas veces usted le haya arrancado un pelo!) No importa cuántos pasos haya dado uno para la formación del propio carácter, dice Edwards, siempre hubo un primer paso, y si el primer paso no fue una decisión controlada por el agente, la responsabilidad del resultado no puede atribuirse al agente. Partiendo de esta base, el segundo paso tendrá las mismas propiedades que el primero, de modo que ese proceso nunca dará por resultado una decisión tomada por un agente responsable de su propio carácter.

Con un razonamiento análogo, podemos demostrar la imposibilidad de la existencia de mamíferos, pues todo mamífero debe ser hijo de una madre mamífera; y si remontamos el árbol genealógico de cualquier aparente mamífero, encontraremos obviamente un ancestro no mamífero cuya progenie, en consecuencia, no podría ser mamífera, y así sucesivamente.[17] Puesto que sabemos perfectamente bien que somos mamíferos, utilizamos el argumento sólo para descubrir la falacia que encierra. Para descalificar el argumento de Edwards, podríamos alegar que no es más que un ejemplo de la estrategia falaz de argumentación que acabamos de mostrar. Pero, en una cuestión tan controvertida como la del yo, esto podría considerarse una petición de principio. Es mejor limitarse a analizar el error o los errores del argumento.

[16] Van Inwagen (1983, pág. 150) reconoce el vínculo: "Debo elegir entre lo enigmático y lo inconcebible. Elijo lo enigmático". Yo, por el contrario, elijo extender los límites de lo concebible.

[17] El ejemplo, con algunas alteraciones, proviene de Sanford 1975. Véase también Sher 1979 y Zaitchick 1977.

Quizás el error consiste en lo siguiente: el argumento presupone que una persona no puede ser plenamente responsable de algo, si ese algo no está enteramente realizado por esa persona. Pero, por supuesto, nada está hecho enteramente por uno, a menos que uno sea Dios; por lo tanto, es ésta una premisa demasiado fuerte para empezar una argumentación. ¿Un artista no puede exigir que se le reconozca plenamente una creación que incorpora, como parte propia, un *objet trouvé*? ¿La explotación ingeniosa de hallazgos casuales le quita el mérito de autor al explotador?

Yo asumo la responsabilidad de lo que hago y de los efectos que mis productos pueden tener sobre el público en general: si mi sopa produce envenenamiento, mi automóvil contamina el aire o mi robot mata a cualquiera, yo, el fabricante, soy el culpable. Y, aunque puedo lograr que mis subcontratistas y proveedores compartan, hasta cierto punto, mis obligaciones, soy el responsable de entregar al público el producto, a pesar de las fallas que tiene. El sentido común dice que la responsabilidad personal se basa en este mismo fundamento lógico. He creado y dejado en libertad a un agente que soy yo mismo: si sus actos ocasionan daño, el responsable es quien los fabrica. Pienso que el sentido común es sabio en este aspecto y debe ser defendido de la sospechosa *reductio* de Edwards que, cabe señalar, no hace un uso esencial de las premisas del determinismo; una *reductio* tan forzada puede construirse también a partir de la hipótesis de un mundo indeterminista.

¿Somos los productos de actos de autocreación? ¿De qué manera? En el capítulo tres he dado algunos ejemplos de las decisiones de metanivel en un autocontrolador: ¿qué rumbo debe tomar el piloto del avión para maximizar la libertad de acción y, en consecuencia, maximizar las oportunidades de mantener el control? ¿El jugador debe optar por una estrategia de juego pautada por estrategias que no siguen ningún patrón de juego? Hay decisiones de metanivel relativamente fáciles de tomar, pues el conjunto de consideraciones aparentemente relevantes es pequeño, sencillo de manejar y visible. Para esta clase de problemas poseemos una solución estándar, la mejor solución, la solución "de libro": mire antes de saltar, una puntada a tiempo ahorra nueve, no llames al lobo, etc. Evolucionamos eventualmente desde esos simples problemas a cuestiones mucho más imponderables.

Estas son algunas de las preguntas que debe responder un complicado agente que se autocontrola: ¿puedo revisar mis proyectos y objetivos básicos con el fin de aumentar la probabilidad de satisfacción? ¿Existen estrategias o políticas extraordinarias, mejores que las mías? ¿Hay un estilo de obrar que se adecue mejor a mis propósitos que mi actual estilo? ¿Mi política actual me conduce

a situaciones con un estrecho margen de maniobra y grandes posibilidades de catástrofe? ¿Cuál sería mi política general respecto del riesgo? ¿Qué clase de agente deseo ser o llegar a ser? (Algunas de estas preguntas podrán tildarse de amorales y ser por consiguiente objetables, pero he mantenido aquí las cuestiones morales deliberadamente al margen, pues es mi propósito demostrar que una gran parte de la estructura del "libre albedrío" ya está fijada antes de la aparición del problema de la moral.) En general, no hay respuestas "de libro" para estas preguntas; como en un juego de ajedrez, uno debe abandonar las aperturas "de libro" a mitad de la partida y lanzarse osadamente al territorio del razonamiento arriesgado y heurístico.

Si desea comprometerse en el proceso de razonamiento de metanivel, el aspirante a un alto grado de autocontrol debe poseer la capacidad de representar sus creencias, deseos, intenciones y políticas como objetos de evaluación. Como señaló Locke, un agente tal tiene el "poder de *suspender* la ejecución y la satisfacción de cualquiera de sus deseos" y, por ende, está en libertad de considerarlos, examinarlos por todos lados y compararlos con otros" (*Ensayo* II, XXI, 118). Pero cuando este examen trasciende la mera resolución de los objetivos en conflicto inmediato (por ejemplo, no puedo ir al teatro y al concierto al mismo tiempo),[18] la comparación puede necesitar una serie indefinidamente larga y no estructurada de consideraciones. Uno debe limitar las propias reflexiones, dar unos pocos saltos a ciegas o posponer la resolución para siempre.

No importa cómo consideremos, evaluemos y reflexionemos, siempre sobrará espacio lógico. Puesto que se trata inevitablemente de un examen limitado e incompleto, es importante que en el proceso de deliberación las consideraciones "correctas" nos sucedan *lo antes posible*. Pero no podemos controlar directamente este orden; no podemos desempeñar el papel de bastonero en el desfile de las consideraciones-a-ser-investigadas, ubicando a cada una en el sitio correspondiente de la fila. En efecto, un control semejante implica un tiempo y esfuerzo, y tenemos mucha prisa. Debemos abandonar el proceso que genera el conjunto incompleto de consideraciones a partir de las cuales actuamos; por ello, siempre nos encontramos, en cierta medida, a merced de dicho proceso. Sin embargo, aun sabiendo que estamos y que estaremos siempre a merced del proceso, podemos dar pasos encaminados a mejorar —si no a garantizar— su

[18] Véase Frankfurt, para las dos diferentes clases de conflicto entre los deseos: conflictos donde los deseos "pertenecen al mismo orden", de modo que si uno de ellos no puede ser satisfecho, el otro se vuelve automáticamente un candidato a satisfacer, y conflictos cuya resolución implica que el agente *rechace* el deseo perdido. Taylor (1976) hace la misma distinción en otros términos: evaluación "cualitativa" y "no cualitativa".

confiabilidad. Como señala Hampshire: "Lo que un hombre *piensa* que hará está estrechamente circunscrito y en ocasiones es predecible casi con exactitud" (Hampshire, 1959, pag. 83). Por consiguiente, tomamos medidas para mejorarnos a nosotros mismos (y ayudar a que otros las tomen) para que en el futuro aumente la probabilidad de pensar las cosas correctas en el momento oportuno.

Pero incluso este proceso de autocrítica reflexiva tiene sus límites. Así como no es aconsejable el rechazo complaciente e irracional de todas las oportunidades de reflexión, tampoco es racional embarcarse en una ronda indefinida de autoexamen. Una vida sin reflexión acaso no es digna de ser vivida, pero una vida que se autoexamina en exceso no es en absoluto recomendable.[19] Además, no existe ningún "libro de respuestas" para la pregunta acerca de cuál debe ser la cantidad justa de moderación.

¿Por qué nos inclinamos a pensar que niveles más altos de reflexión y períodos más largos de autoexamen *tienden* a mejorar el carácter del agente? Supongamos que existen dos agentes A y B nacidos en circunstancias similares y con cualidades similares, sólo que A dedica más tiempo a la autorreflexión y a la consecuente autoelección. ¿Hay alguna razón para suponer que A tenderá a ser un agente mejor que B? Por cierto la hay, si A y B pertenecen a la clase de los seres que aprenden; es decir, si son seres diseñados para propender a la recolección de verdades, pues entonces sus períodos de autoexamen *tenderán* a crear autoconocimiento; ¿y quién dijo que más autoconocimiento es peor que menos autoconocimiento?

Hasta un cierto punto (mal definido), la adquisición de autoconocimiento debe ser una cosa buena, pues, como vimos en el capítulo 3, la anticipación de las limitaciones que podemos encontrar en el futuro debe incluir, eventualmente, las limitaciones que nos son propias.[20] Pero siempre es posible sobrecargar el sistema con un exceso de autoconocimiento. Una característica esencial de esta tarea es la falta de un límite superior objetivo cuya misma superioridad hace más difícil la evaluación de los resultados.

> La evaluación siempre deja espacio para re-evaluaciones. Pero nuestras evaluaciones son las más susceptibles de ser recusadas precisamente en virtud de la profundidad que observamos en el yo. Las evaluaciones profundas son, justamente, las menos claras y articula-

[19] Creo que Kurt Baier hizo una observación al respecto. Ryle (1949) ha observado con agudeza los límites de la reflexión y lo "eternamente penúltimo" de la autoadmonición en "The Systematic Elusiveness of 'I'", págs. 195-198.

[20] Frankfurt (1971) ha subrayado la importancia, para la clase de libre albedrío que podría darnos la responsabilidad, de los deseos de un orden superior: deseos acerca de los deseos que deberían movernos (esto se discute en Dennett 1976).

> das y las más vulnerables a la ilusión y a la distorsión. Aquellas evaluaciones que más se acercan a lo que soy como sujeto —pues, desgajado de ellas, me derrumbaría como persona— son, quizá, las más difíciles de esclarecer. Hay preguntas que pueden ser siempre formuladas: ¿debo volver a evaluar mis evaluaciones básicas?, ¿he comprendido realmente qué es lo fundamental para mi identidad?, ¿en verdad he determinado cuál es el modo más alto de vida? Una vez que nos hemos embarcado en este tipo de re-evaluación, percibimos su carácter peculiar. Es diferente de las evaluaciones menos radicales que se realizan en términos de alguna evaluación básica; por ejemplo, cuando me pregunto a mí mismo si es honesto evadir impuestos o pasar contrabando. Estas últimas evaluaciones pueden efectuarse en un lenguaje que está fuera de discusión. En las preguntas recién mencionadas, el término "honesto" es considerado como irrecusable. Pero en una evaluación radical, los términos esenciales —es decir, aquellos mediante los cuales se pueden efectuar otras re-evaluaciones— son, precisamente, los que están en tela de juicio... ¿Cómo realizar entonces esas re-evaluaciones? No existe, por cierto, ningún metalenguaje en el cual yo pueda juzgar las autointerpretaciones rivales. Si lo hubiera, no se trataría de una re-evaluación radical. (Taylor 1976, págs, 296-297.)

Hemos regresado al vértigo de Nietzsche: la re-evaluación de todos los valores, y ahora parece que el precio a pagar por nuestra capacidad ilimitada de reflexión es el ingreso a los dominios donde la fe en la benéfica tendencia a *perfeccionarnos a nosotros mismos* se vuelve una imposibilidad.[21] Sin duda, el agente que reflexiona sobre sí mismo es más responsable y maduro; esto es, si uno piensa *otorgarle* responsabilidad a alguien para algo —por ejemplo, encomendar nuestros niños a su cuidado— sería prudente elegir el agente más reflexivo. Pero algunos de los virtuosos de la autorreflexión se han extralimitado. Uno de ellos es Nietzsche; el otro, Charles Manson.

La falta de fiabilidad en los niveles más altos se explica fácilmente; los agentes finitos deben usar, entre otros recursos, métodos heurísticos cuando se enfrentan a extensos problemas en un tiempo limitado. Cuanto más extensa es la investigación de un problema,[22] cuanto más grande y desordenado es el pajar en donde se debe

[21] "Vivimos en una nube donde zumban los por qué y los para qué, los propósitos y los razonamientos de nuestras narraciones, de tantas aventuras desastrosas de otros 'yo' análogos al nuestro. Y ese constante girar de posibilidades es precisamente lo que necesitamos para salvarnos de un comportamiento demasiado impulsivo" (Jaynes 1976, pág. 402).

[22] O el espacio de la busca *aparente*. Hofstadter (1983) describe un experimento del tipo "el dilema del prisionero" que él mismo dirigió, en el cual los participantes debían elegir entre "cooperar" o "desertar" en condiciones bastante inusuales. Hofstadter

encontrar la aguja, más arriesgado resultará cualquier método heurístico, con sus inevitables saltos a ciegas.

¿Son esos saltos a ciegas las "elecciones radicales" del existencialismo de Sartre? Taylor arroja luz sobre la pregunta cuando analiza el famoso ejemplo de Sartre (Sartre, 1946) acerca de un joven desgarrado entre dos opciones: cuidar a su madre enferma o unirse a la Resistencia. Se trata de una auténtico problema moral, señala Taylor, por cuanto existen poderosas exigencias morales en favor de ambas alternativas. No se puede declarar simplemente "por elección radical" que la exigencia no atendida no tenía valor. "La verdadera fuerza de la teoría de la elección radical proviene de la intuición de que existen diferentes perspectivas morales, que hay una plenitud de visiones morales ... difíciles de juzgar y, en consecuencia, de elegir" (Taylor, 1976, págs. 293-294).

Existe una pregunta genuina acerca de qué se debe hacer, una pregunta que puede tener una respuesta "objetiva" y "definitiva"; no hay, sin embargo, ningún "proceso de decisión" viable para determinar la respuesta, en el caso de que haya una. Y el tiempo pasa; y debemos actuar. Más tarde, uno se entera retrospectivamente de que tomó la decisión equivocada (equivocada en virtud de consideraciones que uno evalúa ahora y que *evaluaba entonces* y que, quizá, se habrían atendido si se las hubiera explicitado en estos términos en aquel momento). De acuerdo con la lectura de Taylor, Sartre recomendaba el autoconocimiento y el reconocimiento:

> Dado que éste es el predicamento moral del hombre, elegir con lucidez es más honesto, valiente, autoclarividente e implica un modo más alto de vida que ocultar las propias elecciones detrás de la supuesta estructura de las cosas y escapar de la propia responsabilidad a expensas de mentirse a sí mismo y de una profunda duplicidad interior... (Taylor, 1976, pág. 294.)

Pero Taylor afirma que esto no es suficiente. Según la teoría de la elección radical, ¿qué podría impedirnos desestimar nuestras elecciones difíciles de la misma manera que el jugador de ajedrez

---

insiste en que el problema que enfrentaban los participantes tenía una sola solución dictada por la razón: se debía cooperar. Pero ¿cómo debía estar tan seguro de que existía una solución clara, irrefutable, "de libro", en favor de cooperar, cuando muchas personas racionales no lo consideraban así? Ellos vieron aparecer una serie indefinida de consideraciones; hasta donde podían entender, el dilema del prisionero es un *amplificador de incertidumbre*, y por esa razón era dable esperar que provocara un conjunto no uniforme de soluciones por parte del grupo de agentes. Parece que ellos tenían razón y que Hofstadter estaba equivocado. El dilema del prisionero no parece similar al problema de la adición, en el cual cada agente puede confiar en los otros para llegar a la misma respuesta.

desestima sus errores? "Corría el tiempo del reloj y yo di una puñalada en la oscuridad. Resultó ser la puñalada equivocada. Lástima." El jugador de ajedrez reconoce la autoría de sus elecciones (¿cómo podría no hacerlo, a menos que afirmara que está sometido al control remoto de otra persona?); sin embargo, esta clase de autoría es necesaria pero no suficiente para la responsabilidad moral. El fundamento de la responsabilidad, señala Taylor, proviene de la manera peculiar en que llevamos a cabo la re-evaluación radical. No hay, dice Taylor, un metalenguaje posible para tal indagación:

> La re-evaluación se realiza en fórmulas accesibles, pero con un pedido de atención, por decirlo así, a aquello que dichas fórmulas intentan articular... Quien ha luchado con un problema filosófico sabe a ciencia cierta cómo es esta clase de indagación. En filosofía, comenzamos normalmente con una pregunta que sabemos mal formulada desde el principio. Esperamos que sus términos se transformen en el transcurso de la lucha, de modo que al final del proceso respondemos a una pregunta que no podíamos concebir desde el principio.

Cuando la pregunta se refiere a la autoevaluación radical,

> nuestros intentos por formular lo que consideramos importante deben hacer lo posible, como en las descripciones, por ser fieles a alguna cosa. Pero el objeto de esa fidelidad no es un objeto independiente, con un determinado género y grado de evidencia, sino, más bien, una indefinida e inarticulada intuición de aquello que es de importancia *decisiva*. (Taylor 1976, pág. 123.)

Pero si tenemos éxito en el intento, nuestro éxito valdrá más que un acto de autoría, pues nos definiremos a nosotros mismos en el proceso, "articulando" y determinando aquello que era incoado y malformado. Según Taylor, es así como creamos nuestros valores al tiempo que nos creamos a nosotros mismos.

Otros escritores han expresado concepciones similares. En *Fundamentación de la metafísica de las costumbres*, Kant afirma que las leyes que *nos damos a nosotros mismos* no nos obligan. Esto sugiere que el yo que resulta de ese proceso no está constreñido por las leyes que promulgamos, pues el yo está constituido (en parte) por esas mismas leyes y (en parte) creado por un *fiat* que articula y define algo que hasta el momento estaba hecho a medias o era informe. Mill describe la resolución de una crisis moral en un lenguaje que, en otros escritores, habría suscitado tesis metafísicas extravagantes:

> Pero es evidente que "yo" soy las dos partes de la confrontación; el conflicto está entre yo mismo y yo; entre (por ejemplo) mi deseo de un

placer y mi temido autorreproche; ¿qué es lo que me mueve a Mí o, si se prefiere, a mi Voluntad, a identificarme como una parte y no, más bien, con la otra? La razón es que uno de mis "yo" representa un estado de mis sentimientos más permanente que el otro. Después de haber cedido a la tentación, se terminará el "yo" que desea, pero el "yo" que reprocha durará hasta el fin de mi vida. (Mill 1867, pág. 452)

Nozick encuentra aun otra analogía para esos momentos acumulativos de autodefinición: la medición u observación que en mecánica cuántica "'derrumba el paquete de ondas', reduciendo la superposición [de diferentes estados posibles] a un estado particular..." (Nozick 1981, pág. 298). Esto puede resultar una atractiva analogía para muchos, pero su valor es *dudoso*. Por esta razón, el derrumbe del paquete de ondas descrito en términos de mecánica cuántica es evidentemente difícil de comprender; de modo que corremos el riesgo de explicar un "misterio" por medio de otro más oscuro. (Los filósofos con frecuencia disfrutan al cometer este pecado en el acto mismo de la acusación: ésta es la falacia del *obscurum per obscuris*.) Además, la analogía invita a que se la tome muy en serio, como una explicación de la microfísica última de la elección humana, si bien el mismo Nozick se ha ocupado de anticiparse a ese tipo de lectura.[23]

Esta concepción de la autoformación presenta el gradual desarrollo del yo como un proceso sustancialmente diferente de otros procesos de transformación material que podrían servir de metáforas apropiadas para estos *hyléfobos*[24] con poca imaginación: la formación de cristales, las erosiones, la maduración, la formación de tejido cicatricial, etc. La autoformación, a diferencia de estos procesos, es sensible a la información, es pasible de ser conducida a niveles infinitos de crítica de metanivel y creativa como lo es el arte: las formas que surgen contribuyen a constituir los cánones por las cuales se las juzga.

Pero esta posición, a pesar de su poder persuasivo, sigue tan vulnerable como siempre a la recusación escéptica de Edwards: si me he definido a mí mismo de tal manera que puedo, en ocasiones, escuchar la voz de la razón, ello se debe a mi buena suerte; y si otra persona es lo bastante desafortunada como para comenzar su vida

[23] "El propósito de esta comparación no es hacer derivar el libre albedrío de la mecánica cuántica ni usar una teoría física para probar que el libre albedrío existe, ni siquiera decir que el no-determinismo en el nivel cuántico permite suponer el libre albedrío. Más bien, queremos ver si la teoría cuántica brinda una analogía ..." (pág. 298). Véase también su discusión del yo como "una artefacto de autosíntesis", págs. 105-114.

[24] *Hylefobia* es un término inventado por mí que se aplica al miedo al materialismo. Discuto sobre la hylefobia humanística en Dennett 1978b, pág. 252.

con un carácter lastrado por defectos ancestrales, sus delitos serán igualmente imputados a la mala suerte. Ninguno de los dos fue responsable de ser lo que era el día de su nacimiento, y ninguno de los dos será, o podría ser, responsable de lo que haga con posterioridad, mientras que nuestro carácter de hoy sea el resultado (no importa si constructivo o creativo) determinista de nuestro carácter de ayer, que a su vez es el resultado determinista de nuestro carácter de anteayer, etcétera.

El hecho de que algunos nazcan con bastante talento artístico para desarrollar "buenos caracteres", ¿se debe a la "suerte"? El concepto de suerte y el papel que desempeña en los argumentos sobre el libre albedrío ha recibido, sorprendentemente, poca atención por parte de los filósofos.[25]

3. Probar suerte

> *Y cuántos hay que, habiendo llevado*
> *una vida intachable, son solamente afortunados*
> *por haber escapado así a muchas tentaciones.*
> —Kant, *Metafísica de las costumbres*

La suerte es un concepto curioso. Cuando decimos que algo ocurrido fue mera suerte, deberíamos preguntarnos cuál es el concepto que se opone al de suerte. Creer que "existe la suerte" —algo que puede traer la pata de un conejo— es una superstición, pero hay una manera de calificar los acontecimientos y las propiedades como *meramente* afortunados que no es supersticiosa ni engañosa. La diferencia entre ambos conceptos se hará manifiesta en el siguiente experimento mental.

Supongamos que los Estados Unidos y Rusia se han puesto de acuerdo para resolver de ahora en adelante todas sus desavenencias jugando a "cara o cruz" en Ginebra. Ambos países designarán en consecuencia un participante para el gran certamen internacional de cara o cruz. ¿Quién representará a los Estados Unidos? Espero que alguien tenga la brillante idea, grandiosa y a la vez democrática, de organizar un concurso nacional de cara o cruz en el cual participen hombres, mujeres y niños que serán sistemáticamente eliminados. El ganador será, por cierto, el participante ideal para enviar a Ginebra, ¡el ciudadano más afortunado de los Estados

[25] Nagel (1979) y Williams (1981) han llamado la atención de los filósofos sobre el curioso papel que desempeña la suerte en las cuestiones de responsabilidad moral, pero ambos dan por sentado el concepto de suerte.

Unidos!, ¡alguien que ha tirado la moneda 28 veces consecutivas sin perder ni una sola vez! ¿Esa persona tiene realmente más probabilidad de ganar en Ginebra que otro ciudadano escogido al azar? Si la suerte existiera, es decir, si fuese una propiedad manifiesta de las personas y de las cosas, entonces el imaginario certamen funcionaría como una especie de detector o amplificador. Pero, a menos que estemos muy equivocados, la suerte no es así. La suerte es *mera* suerte y no un don auténtico, manifiesto.

El ganador del certamen imaginario es afortunado (en el sentido no supersticioso) y, en consecuencia, debe decir: "pienso que fui afortunado", pero de ninguna manera: "pienso que soy una persona afortunada", pues no existe tal cosa; no se puede ser afortunado en general, sino afortunado sólo en ocasiones. Pero al ganador le resultará difícil no sentirse un elegido, alguien especial, tocado mágicamente por la buena suerte. Lo mismo les ocurre a los sobrevivientes de una catástrofe (naufragios o accidentes en las minas); normalmente sacan conclusiones extravagantes del hecho de *haber sobrevivido* mientras otros perecieron. También el ganador del certamen caerá en la tentación de creer que la suerte es algo real, una fuerza de la naturaleza que le es favorable.

Incluso podría sentirse *responsable* de haber ganado el certamen, como si hubiese hecho realmente algo para merecer tanta atención. Supongamos ahora (sólo para aumentar la ilusión) que el ganador ignora la existencia del certamen. Hasta donde sabe, ha sido invitado a tirar la moneda desde hace unas pocas semanas y ha obtenido una serie increíble de aciertos (ganó 28 veces consecutivas). En este caso, el ganador se preguntará si la moneda no está trucada, o bien se volverá religiosamente supersticioso.[26] Ciertamente, los

[26] Este experimento mental está inspirado en uno de los clásicos juegos de apuestas más elegantes y traicioneros; como ignoro su nombre —probablemente ya debe tener uno— lo llamaré *la pirámide del pasador de datos*. Usted consigue una lista de apostadores serios, la divide por la mitad y envía esa mitad con el dato de que el equipo A ganará el campeonato la semana próxima, y la otra mitad, con el dato de que el equipo A perderá. Luego de una semana, los apostadores que pertenecen a una de las dos mitades de la lista habrán recibido de su parte un dato correcto, libre de cargo. Descarte la otra mitad de la lista. Divida la que queda otra vez por la mitad y mande su segunda tanda de datos; si bien este procedimiento disminuye la clientela de incautos, aumenta la fe en su clarividencia, pues ya lleva en su favor dos aciertos. Después de unos pocos "exitos" más, anuncie que se ha terminado el período de prueba gratuito y que de ahora en adelante se debe pagar. Como diría Skinner, la belleza del esquema es que uno no tiene que atenerse a un grupo que disminuye tan rápidamente. Algunas personas son muy incautas respecto de "los refuerzos suplementarios del azar". La próxima vez que envíe datos, ofrezca un "descuento" a aquellos para quienes su proporción de éxitos no es tan brillante (digamos, una de cada cuatro predicciones ha sido falsa) y encontrará seguramente muchos apostadores.

que estamos al tanto no nos sentiremos atraídos por esa línea de pensamiento. Sabemos que todo certamen tiene inexorablemente un ganador que sólo lo es mientras dura el certamen.

Ahora bien, ¿existe en aquellos de nosotros que se sienten comandantes del propio destino, una ilusión de responsabilidad personal o de poseer cierta excelencia, semejante a la ilusión que experimenta el ganador de nuestro certamen? Seguramente. Cualquiera que se acredite el éxito de haber ganado el certamen y anticipe futuros éxitos sobre la base de la experiencia pasada, está cometiendo un error. ¿No estaríamos acaso cometiendo la misma equivocación cuando pensamos que nosotros —por lo menos, algunos de nosotros— somos auténticamente responsables de nuestros actos?[27]

En un sentido, todos somos extraordinariamente afortunados por el hecho de estar aquí, en el círculo de los ganadores. ¡Ninguno de *nuestros* antepasados tuvo la desgracia de morir en la infancia! Cuando pensamos en los millones de predecesores que quedaron fuera de juego en las primeras etapas del certamen de selección natural, ¡concluimos que las apuestas en contra de la posibilidad de *nuestra* existencia deben haber sido astronómicas![28]

La diferencia fundamental entre nosotros y el ganador del certamen es que no podemos acreditarnos personalmente el éxito de nuestros antepasados: sólo nuestros genes lo pueden hacer. El certamen que ganaron prueba que poseían verdadera excelencia. Hay que ser bueno en algo (en realidad, en todas las cosas en general) para participar en el certamen que hoy se está jugando. La persona que gana un concurso de preselección en un juego de destreza probablemente jugará mucho mejor que la persona elegida al azar, en los concursos siguientes; la destreza, a diferencia de la suerte, es predecible. Y, puesto que la capacidad de autocontrol y de deliberación fue puesta a prueba a lo largo de millones de años, ¿existe una base real que sustente nuestras grandes expectativas respecto de la capacidad de deliberación —y, en especial, de autocontrol— de nuestros compañeros, los seres humanos? Si no estuviéramos bien pertrechados en ese rubro, no habríamos salido airosos en esta etapa del certamen. Ciertamente, existen personas desafortunadas por causas ajenas a su voluntad, aunque hayan nacido de padres que se autocontrolan muy bien. Ellos están excu-

[27] Véase la discusión sobre este tema con Hospers 1961 y con Hook 1961b.

[28] Y hablando de astronomía, pensemos que hemos tenido la suerte de que nuestro planeta esté compuesto de elementos pesados, que son muy raros en el universo. ¡De modo que tenemos algo cómodo y sólido en donde pararnos! No sé si somos o no afortunados por el hecho de estar aquí, pero, puesto que *estamos* aquí, podemos tener la capacidad de descubrir hechos sorprendentes sobre las "leyes de la naturaleza". Véase Gale 1981.

sados. No los consideramos responsables. Pero poseemos sobradas razones para esperar mucho del resto. ¡No solamente somos afortunados sino que, además, somos hábiles!

La relación entre habilidad y suerte es compleja. Según una máxima muy común entre los atletas, la suerte tiende a promediarse y los mejores jugadores son aquellos bien equipados para capitalizar las oportunidades afortunadas. Cuanto mejor es uno, menos necesita de la suerte y, en menor medida, los éxitos son meramente afortunados. ¿Por qué? Porque cuanto mejor es uno, más control tiene sobre el propio desempeño.[29] Cuando el astro de baloncesto encesta la pelota diez veces consecutivas en un partido, no decimos que tuvo suerte; pero si soy yo el que encesta diez veces consecutivas la pelota, entonces sí dirían que tuve suerte.[30] El astro de baloncesto no es tan bueno como para prescindir por completo de la suerte, pero en su caso, el umbral de lo que consideramos suerte está colocado mucho más alto. Esperamos más de él. De igual modo, si alguien tiene la habilidad suficiente para perfeccionarse a sí mismo, el éxito que logre en este aspecto no es imputable, en cada ocasión, a la suerte sino a su habilidad. Esperamos más de él. Como dice el refrán: *Noblesse oblige.*

Pero aun así, podríamos alegar que en las circunstancias de un agente existen dos factores meramente imputables a la suerte: la fuerza inicial, el talento o carácter con los que ha tenido la suerte de nacer, y las coyunturas felices que ha encontrado durante el período de autocreación. De manera similar, estos factores tienden a impugnar la pretensión del agente de que se le reconozca la responsabilidad personal de su propio carácter. Analizaremos estos factores por separado.

Supongamos —aparentemente, es cierto— que las personas nacen con cualidades cognitivas y rasgos potenciales de carácter notablemente distintos. Por ejemplo, hay personas que provienen de una larga línea de antepasados brillantes (o apasionados, o bien musculosos) y parecen poseer cualidades iniciales muy diferentes de las de sus contemporáneos. ¿Es esto "terriblemente injusto" —como dice Williams (1981, pág. 228)— o está destinado a provocar algo terriblemente injusto? No necesariamente.

Imaginemos una carrera con una línea de salida que varía según los competidores. Los nacidos en enero se colocan un metro

[29] Alex Bird (un famoso inglés que hizo una fortuna en las carreras de caballos) dijo, en cierta ocasión: "Nunca me consideré una persona afortunada. Soy un cobarde, por eso no puedo ser jugador. Pero trabajo duro. Cuanto más duro trabajo, más afortunado me vuelvo" (*Sunday Observer*, Londres, 24 de abril de 1983).

[30] Véase Winisatt 1980, sobre la relación entre habilidad, oportunidad y tolerancia al error.

más adelante que los nacidos en febrero y 11 metros más adelante que los nacidos en diciembre. Seguramente, nadie puede evitar el hecho de nacer en un mes o en otro. ¿No es esto palmariamente injusto? Sí, si se trata de una carrera de cien metros llanos; no, si se trata de un maratón. En un maratón, esa pequeña desventaja no tiene la menor importancia, pues podemos confiar en que otros acontecimientos fortuitos produzcan incluso mejores efectos. En un largo maratón, a los mejores corredores se les concede una considerable ventaja en la salida, y nunca escuché que nadie se quejara al respecto. El buen corredor que comienza en la retaguardia, si es tan bueno como para merecer el premio, tendrá innumerables oportunidades de superar la desventaja inicial. A pesar de la amplia aceptación de la que goza, muchos pueden dudar de la justicia del procedimiento. Estarán de acuerdo, entonces, en que el arbitrario sistema que considera el mes de nacimiento es más justo. ¿Es lo bastante justo para no tener que preocuparnos por él? Por supuesto; después de todo, la suerte a la larga se promedia.

¿Tenemos alguna razón para creer que el proceso de desarrollo moral o la adquisición de la facultad de ser agentes se parece más a un maratón que a una carrera de 100 metros llanos? Sí, y tenemos abundantes pruebas al respecto.

Por este motivo, el desarrollo moral no es una carrera con un único ganador y el resto de los competidores alineados detrás de él, sino un proceso que tarde o temprano conduce a las personas a una especie de meseta del desarrollo, y que no es muy distinto del proceso de aprendizaje de la lengua nativa. Algunos llegan fácil y rápidamente a la meseta, en tanto que otros necesitan esfuerzos compensatorios para superar las desventajas iniciales.[31]

Pero, en cierta manera, todos provenimos de la misma sociedad. Cuando a las personas se las juzga "bastante buenas", su educación moral termina; así, con la excepción de los discapacitados (los retardados o los psicópatas, por ejemplo) puede decirse que la ciudadanía está compuesta por individuos cuyos talentos son más o menos equivalentes en lo que respecta a las demandas de dicha ciudadanía. Comúnmente *se cree* que tanto las diferencias iniciales como sus consecuencias más o menos afortunadas se promedian. ¿Se trata de un grave error del sentido común? Podemos plantear nuevamente la pregunta en el terreno de los deportes con el propósito de encontrar mejores políticas a seguir.

[31] Véase Sher 1979. Véase también Hampshire 1959, pág. 186: "Los seres humanos pueden identificar sus propias limitaciones como un aspecto del mundo que será autoconscientemente aceptado o, si es posible, deliberadamente modificado, pero no simplemente acatado."

Imaginemos ahora que intentamos cambiar las reglas del baloncesto de la siguiente forma: si el árbitro decide que el encestamiento se debió exclusivamente a la suerte, no se cuentan los tantos; si nota que la mala suerte se está ensañando con uno de los equipos, le concede privilegios compensatorios. En los deportes, aunque uno planee las jugadas y haga lo posible por ganar, se acepta la suerte y se la tiene en cuenta. La suerte, después de todo, es la misma para todos; nadie tiene *más suerte* que otro, aunque *haya sido* lo bastante afortunado como para haber nacido con más talento. Pero eso también es justo, pensamos. No creemos que el único certamen justo es aquel cuyos contrincantes tienen exactamente las mismas cualidades. La fuerza de uno puede vencer la astucia del otro y viceversa. Todo lo que pedimos es que sean parejos en valentía y, si en determinada ocasión la fuerza de uno vence al otro, eso será"muy malo" para este último, pero de ninguna manera será injusto.

Cuando las personas de talento triunfan, no es porque tienen suerte sino porque tienen talento. [32] ¡Ah, pero, en primer lugar, son afortunadas por tener talento! Algunos intérpretes no nacieron sino que se hicieron talentosos. Durante años, practicaron muchas horas por día hasta lograr la excelencia y en el transcurso del proceso se negaron a muchos placeres y oportunidades que nosotros, intérpretes menos diestros, no habríamos vacilado en aprovechar. ¡Ah, pero ellos tuvieron la suerte de nacer con ese empuje y esa iniciativa que les permitieron desarrollar su técnica y adiestrarse a sí mismos! No siempre. Hay quien no nació con ese temperamento sino que lo aprendió de un entrenador o de un sabio maestro. ¡Pero tienen la suerte de poseer la inteligencia necesaria para comprender las lecciones del buen entrenador! (Después de todo, no todos los jugadores son capaces de aprovechar su enseñanza.) Y, obviamente, tienen la suerte de vivir en la misma ciudad donde vive el entrenador que hace milagros. Más aun, ellos tienen la suerte de no haber nacido ciegos y de que no los partiera un rayo cuando se dirigían a la escuela. En realidad, como subrayé antes, ¡somos astronómicamente afortunados por el mero hecho de haber nacido!

Como lo demuestra este pequeño diálogo exasperante, hay una tendencia a considerar "lo afortunado" y "lo desafortunado" como

[32] Supongamos que el ganador del certamen de cara o cruz recibe un millón de dólares de premio. Podemos decir entonces que no sólo fue afortunado al ganar sino afortunado por tener un millón de dólares (que no ganó trabajando). Pero de ahora en adelante, ya no es afortunado: es rico. Si usted es rico, ¡no necesita ser afortunado! Si la suerte fuera algo parecido a un don, sería verdaderamente injusto dar un millón de dólares al ganador que ya ha probado que se encuentra en una posición muy ventajosa en el mundo: tiene suerte.

términos complementarios y exhaustivos, que no dejan espacio para la habilidad. Según esta perspectiva, nada, en principio, puede ser considerado como una habilidad o como el producto de una habilidad. Esto es un error. Una vez que se reconoce que entre el éxito afortunado y el fracaso desafortunado hay libertad de acción para la habilidad, se evapora el argumento que parece demostrar que nadie es responsable. La suerte se promedia y la habilidad es la que tiene la última palabra. La suerte desempeña en ocasiones un importante papel, por cierto, pero quien crea que todas las pérdidas se explican, *en última instancia*, por la mala suerte y todas las victorias, por la buena suerte, está abusando del concepto de suerte.

Esto se ve claramente en el terreno de los deportes que utilicé para mis ejemplos. Pero, ¿qué sucede cuando trasladamos el tema en discusión del dominio de los deportes al más serio y grave de las decisiones morales? ¿Nos seguirán sirviendo las intuiciones sobre la suerte que parecían tan acertadas? El análisis de la suerte y de la destreza en los deportes fue sólo un precalentamiento y no el tema principal. Así, dejando de lado el precalentamiento, ¿qué deberíamos decir del pobre B del ejemplo de Edwards, descrito como alguien que no tuvo la suerte de nacer con el coraje y la fuerza de carácter necesarios para soportar la cura?

Muchos autores —por ejemplo, Nagel en "Moral Luck" (1971)— consideran necesario excusar a las personas como B, pues sólo la mala suerte explica el fracaso de B. En ocasiones, esto es así, pero ¿por qué habría de serlo siempre? No conocemos a B lo suficientemente como para saber si su problema se atribuye a la mala suerte o a su fracaso ante una tarea con respecto a la cual teníamos todas las razones para creer que cumpliría con éxito. ¿Es B semejante al campeón que "erró" un tiro fácil y que vivirá abatido durante años hasta olvidar el error o al jugador que, luego de intentar una jugada desesperada, y de perderla, se dirige a los vestuarios sintiéndose como un héroe sólo por haberse esforzado a tal extremo? ¿Por qué razón las distinciones que encontramos con tanta facilidad en el campo de los deportes no resultan aplicables al más vasto escenario del mundo?

Porque —ésta es la respuesta— cuando decimos que "tenemos razones para creer" que un agente hará algo bien, no nos estamos expresando correctamente. Si el determinismo es verdadero, es sólo porque nunca tenemos toda la información que es, en principio, relevante para esas cuestiones en las cuales esperamos que alguien haga algo distinto de lo que, en realidad, termina haciendo. Si conociéramos los mínimos detalles, podríamos darnos cuenta de que la persona "no podía evitar" hacer lo que efectivamente hizo en esa

ocasión. En los capítulos cinco y seis, mostraremos que esta línea de argumentación es errónea. (Para algunos, esto constituye el núcleo mismo del problema del libre albedrío, y habrán estado aguardando con creciente impaciencia que yo lo plantee. Espero mostrarles que la cuestión no puede ser planteada correctamente sin estas digresiones preliminares.)

Todos somos seres "dotados". Es decir, sostengo que yo y mis lectores somos miembros de la comunidad de los que dan y consideran razones. Todos estamos dotados de la capacidad de deliberar; de lo contrario, no estaríamos aquí, en el mundo adulto de los lectores y los libros. Pero faltan algunos de nuestros parientes y amigos. Unos son niños; otros, infantiles o incapaces de razón por motivos ajenos a su voluntad. Unos pocos *pueden* ser culpables de su desgracia; por ejemplo, los que destruyen a sabiendas su cerebro abusando de las drogas o del alcohol; pero, una vez que se han incapacitado a sí mismos, están excusados de toda responsabilidad. Tomamos medidas para mantenerlos alejados de aquellas situaciones en las cuales podrían dañarse gravemente a sí mismos o dañar a los demás.

Nosotros, los dotados, somos muy buenos para deliberar y para controlarnos a nosotros mismos y, por consiguiente, esperamos mucho de los demás en ese aspecto. Sobre la base de estas expectativas, colocamos a los dotados en posiciones de confianza y de responsabilidad y luego *contamos* con que *ellos* harán lo correcto. Y, puesto que son dotados, no atribuimos a la suerte su buen desempeño. Cuanto mejores son, menos participa la suerte en el resultado de sus deliberaciones.

No somos totalmente responsables de ser responsables, de la misma manera que el campeón de un deporte no es totalmente invulnerable a la suerte. Después de todo, mañana podría aparecer un asaltante en alguna callejuela y romperme los sesos. Entonces, tendría la mala fortuna de perder mi condición de ciudadano responsable, por culpas y errores que me son ajenos. Pero, si bien en este último caso sería víctima de la mala suerte, el hecho de que hasta el momento no me haya topado con ningún asaltante no es sólo una cuestión de suerte; en efecto, soy muy hábil para evitar problemas. En realidad, nosotros, los agentes responsables, estamos diseñados para tender a evitar las situaciones en las cuales podríamos perder nuestra excelencia y, en consecuencia, nuestro *status* de ciudadanos responsables. Cuidamos de nosotros mismos. Tratamos de conservar sistemáticamente nuestro *status*; pero, lógicamente, a veces sobrevienen tiempos difíciles. Lo inesperado e inevitable sucede. Y, entonces, somos desafortunados.

Podemos imaginar (con cierta dificultad, creo) un ser que ha pasado su vida tomando decisiones"correctas" gracias a la buena suerte, pero nunca gracias a su habilidad o destreza. En cada vuelta de la vida, le ha sucedido escoger la dirección correcta, salvándose así de malhechores, calumnias y desgracias. Un desafortunado traspié, y ¡zas! queda condenado por el resto de su vida al infortunio. Pero nosotros no somos así.

"Allá, pero por la gracia de Dios, me encamino" (John Bradford, ¿1510?-1555). Un pensamiento muy curioso que se repite con frecuencia en nuestros días con un significado aun más curioso: podría ser verdaderamente desafortunado; podría sin embargo sufrir una gran desgracia. No obstante, sería altamente improbable que una serie de circunstancias accidentales *me* convirtieran en, digamos, una asesino. En cambio, es probable que los acontecimientos normales y cotidianos sirvan de detonante para el joven rufián que hasta el momento se ha salvado por un pelo de meterse en graves problemas. El día en que alguien cometa la equivocación de insultarlo, ese día comenzará su carrera de criminal violento. Si la clase de tentación que lo convierte en asesino apareciera ante mis ojos, pienso, casi con absoluta certeza, que la resistiría. Si no pudiera resistirla, sería yo el único culpable. Se supone que soy bueno para resistir a semejantes tentaciones. Tanto es así, que *ni siquiera las considero como alternativas posibles.* Piénsese en los buenos jugadores de ajedrez, que jamás reparan en las oportunidades de realizar una jugada tonta. (Véase Hofstadter 1979, págs. 236 y sigs., sobre el "descarte implícito".) No es solamente la suerte lo que les impide hacer jugadas "tontas".

Si el rufián es un "tonto" en la vida (y lo es), esto es algo muy malo para él y acaso deba imputarse a su mala suerte el hecho de que sea tan poco hábil para tomar decisiones. Sin embargo, hacemos algunas concesiones. Para los "tontos", el umbral de expectativas lo colocamos muy bajo. Por ejemplo, mantenemos a los niños alejados de las situaciones en las que sus juveniles poderes de deliberación podrían conducirlos a errores terribles y lamentables. Lo que es suerte para los niños no lo es para los adultos.

Todo esto está muy bien, pero —se me objetará— ¿qué pasa con aquellos seres desafortunados que parecen adultos, se mueven en el campo de juego de los adultos y se meten en situaciones difíciles en las cuales toman decisiones extremadamente lamentables? Si podemos identificarlos, debemos hacerlo. Y, luego, debemos tratarlos como niños y excusarlos. Pero, entonces, ¿quedaría alguien a quien inculpar? Aun cuando algunas de estas personas estén dispuestas a culparse a sí mismas, ¿harán bien en inculparse? Cuando me culpo a mí mismo de algún error que no puedo atribuir

a la suerte, ¿por qué le concedo a mi *yo* la posesión de un dominio tan amplio? Me ocuparé de esta cuestión en los tres últimos capítulos.

4. Resumen

Hasta el momento, no hemos resuelto nada en forma definitiva. Las partes de la estatua siguen aún"sin terminar". En el capítulo dos, esbozamos una descripción parcial de nuestra racionalidad y, en los capítulos 3 y 4, retomamos los puntos que quedaron pendientes. Aún subsisten problemas relativos a la racionalidad, de los cuales nos ocuparemos en el capítulo 5. En el capítulo 3, realizamos un somero análisis sobre el control y el autocontrol, que fue profundizado en el capítulo 4 y que puliremos cuidadosamente en los capítulos 5 y 6. En este capítulo, presentamos una visión naturalizada del proceso gradual de autodefinición y de nuestra aparición como agentes, agregando detalles a la visión evolutiva que expusimos en el capítulo 2. Vimos (de una manera confusa) la presencia del azar y de la suerte en distintas áreas y decidimos trabajar un poco más estas superficies de la estatua en los capítulos 5 y 6.

El temor que enfocamos en este capítulo es el temor de que ninguna teoría naturalista pueda definir el yo distinguiéndolo de una simple ficha de dominó dispuesta en una cadena de fichas. No queremos ser meras fichas de dominó sino agentes morales. Revisemos ahora qué características especiales encontramos en nuestro yo, concebido de una manera naturalista. Solamente algunas porciones del universo físico tienen la propiedad de estar diseñadas para resistir a la propia disolución o para emprender campañas locales contra la tendencia inexorable de la Segunda Ley de la Termodinámica. Sólo algunas de estas porciones tienen la propiedad adicional de ser inducidas a tener expectativas confiables respecto de lo que ocurrirá en el futuro inmediato y, por consiguiente, a poseer la capacidad de controlar las cosas y de controlarse a sí mismas; y sólo algunas de estas últimas tienen la facultad más alta de autoperfeccionarse significativamente (mediante el aprendizaje). Unas poquísimas tienen la capacidad ilimitada (que requiere un lenguaje de autodescripción) de "autoevaluación radical". Estas porciones del mundo son, en consecuencia, lugares de autocontrol, de talento y de decisión. Tienen proyectos, intereses y valores que crean durante el proceso de autoevaluación y de autodefinición. ¿Podría haber una porción del mundo menos parecida a una ficha de dominó?

5

Actuar según la idea de libertad

No existe el tiempo sin desasosiego;
no existe para los mudos animales que viven
sin inquietud.
—Kierkegaard

Y una de las funciones más generales y
profundas de los organismos vivientes,
es la de mirar hacia adelante y producir
futuro, como dijo
Paul Valery.
—Francis Jacob (1982, pág. 66)

1. Cómo seguir deliberando en los tiempos que corren

Nuestras deliberaciones nos importan mucho y nos desagradaría enterarnos de que, en ese aspecto, somos víctimas de un engaño. Planeamos el futuro, yacemos en vela royendo el hueso de la indecisión, preguntándonos el porqué y el cómo de nuestros actos y prometiéndonos obrar con más circunspección en el futuro. Cuando integramos un jurado, analizamos cuidadosamente las pruebas con el fin de dar un veredicto responsable. ¿Quizá se trata de una preocupación y un empeño desperdiciados? ¿Hay en ello cierta ilusión o impostura? Para los que temen que el determinismo sea verdadero, la hay.

Comenzaremos por una visión omnisciente del universo. El entramado de la causalidad se extiende ante nosotros junto con la línea del tiempo, desde el alba de la creación (a la izquierda) hasta la extinción de todo calor en el universo (a la derecha); y suponemos, como Laplace, que su historia está determinada. En el tiempo *t*,

exactamente en el punto medio, se encuentra Alicia en vías de decidir "si mañana irá a Londres". *Observamos* que en *t* no hay ramificaciones hacia el futuro; si las hubiera, Alicia aún no podría decidir ir o no mañana a Londres en *t*. Pero vemos que aun cuando no haya decidido en el tiempo *t* no ir mañana a Londres, su decisión ya está determinada. Si Alicia piensa en *t*, que la posibilidad de ir mañana a Londres es una posibilidad genuina, una opción "viva" cuya decisión pondrá en práctica o bien anulará, se equivoca por completo. Piensa en esa forma porque su perspectiva es limitada e insuficientes sus concimientos.

Si pudiera ver el mundo desde nuestra posición ventajosa, sabría con certeza que no existe ni existió nunca la posibilidad de ir mañana a Londres. La "posibilidad" que vislumbró en el tiempo *t* no fue sino producto de su imaginación al apoderarse del espacio que la ignorancia dejó en blanco. Desde su perspectiva, el pasado es un todo inmutable y el futuro, "abierto" y rico en oportunidades y posibilidades. Desde nuestra perspectiva imaginaria, su futuro es tan unificado y fijo como su pasado. Tampoco existe un verdadero "ahora" que separe el vasto futuro de la delgada línea del pasado.

Si el determinismo es verdadero, ¿será correcta nuestra perspectiva y equivocada la de Alicia? El hecho de que ni nosotros ni Alicia podamos adoptar el punto de vista omnisciente respecto del futuro, no implica que la visión desde esta inaccesible perspectiva sea por ello menos cierta. Si el determinismo es verdadero, creer que hay más de un futuro posible será una inferencia aparentemente falsa; sólo lo real es posible y todo lo que nos ocurra será la única posibilidad que el futuro nos tenía reservada. Ahora bien, como toda deliberación presupone la existencia de múltiples posibilidades, el determinismo y la deliberación serían, en consecuencia, incompatibles.

Pero, ¿hasta dónde llega la supuesta incompatibilidad? Si hemos de hacer justicia a la tesis, deberíamos desembarazarnos, en primer término, de todas las ideas erróneas que se confunden con la mejor versión de la tesis; así, comprobaríamos que aun la mejor versión es errónea.

Primer paso: se dice con frecuencia que si el determinismo es verdadero, toda deliberación será imposible. Pero esto tiene que ser falso, pues las personas deliberan a diario y, además, lo que es real es posible. Sin embargo, la realidad de las deliberaciones cotidianas no nos demuestra que el determinismo sea falso (de otro modo, la física sería una ciencia mucho más simple).

Segundo paso: si el determinismo es verdadero, las deliberaciones no serán *efectivas*, pues "no establecerán ninguna diferencia". Pero también se comete un error al pensar que "establecer una

diferencia" equivale a "ser eficaz". En este mismo capítulo investigaremos con más detalle el significado de "establecer una diferencia".

Aun en un mundo completamente determinado existen suficientes razones para distinguir las deliberaciones que contribuyen sustancialmente a la causación de un acontecimiento de las que no conducen a resultados relevantes. El preso que pasa los días y las noches maquinando su fuga y muere en la cárcel se embarca en deliberaciones cuyos únicos efectos son, por así decirlo, un mayor número de arrugas en la frente y un aumento de la presión sanguínea. Pero su compañero de celda, que también participa en esas deliberaciones, se decide por un plan, lo pone en práctica y escapa de la prisión. Si "establecer una diferencia" en un mundo determinista significa esto, entonces hay deliberaciones que "establecen una gran diferencia" y otras que no lo hacen. Se podría aducir, por ejemplo, que en los casos de asesinato premeditado la víctima estaría aún viva, si no hubiera existido premeditación.

Tercer paso: aunque una deliberación determinada contribuya con eficacia a una cadena causal, no será nunca una *verdadera* deliberación pues su resultado estará tan determinado como su origen. Pero ¿por qué lo determinado debe comportar siempre una falsedad o una ilusión? (¿Una tormenta determinada no sería acaso una verdadera tormenta? ¿Un accidente de tránsito determinado no sería acaso un accidente de tránsito real?) Declarar, sin fundamento, que una deliberación determinada no es una deliberación real es una petición de principio. Sin embargo —he aquí la réplica—, en toda verdadera deliberación existe una *oportunidad* genuina, con dos caminos "abiertos al agente". Su deliberación clausura uno de ellos y escoge el otro. Si el resultado de la deliberación fuera en sí mismo determinado, estaría "siempre" determinado; es decir, sólo habría tenido una oportunidad aparente desde el principio.

El tercer paso presenta una objeción que vale la pena mencionar, pues se trata de una objeción clara, familiar, irresistible y completamente falaz. En este capítulo demostraré que el argumento descansa en un cúmulo de malentendidos acerca de la naturaleza de la deliberación, de lo que es —o podría ser— una oportunidad, de lo que implica evitar algo y, en consecuencia, del significado del término "inevitable". Pero antes de ocuparme del argumento y del diagnóstico de sus errores, quiero que se advierta la naturaleza tan peculiar e inestable de su conclusión.

Supongamos que alguien que cree en el determinismo acepta la conclusión del argumento. Creerá, entonces, en la imposibilidad de toda deliberación *real* y de oportunidades futuras *reales*. Pero, ¿cuál debería ser su actitud al tomar conciencia de esto? ¿La resig-

nación, la apatía? Obviamente, si el determinismo es verdadero, la actitud que adopte —cualquiera que sea ésta— está predeterminada. No obstante, puede que no esté determinada a la manera de la *sphex*, sino por la consideración de las razones más válidas. De todas formas, ¿cuál sería la actitud *correcta* a adoptar? En general se supone que si la persona es racional, se verá obligada a abandonar toda deliberación porque, según se ha demostrado, es un *modus operandi* inútil.

Nunca puedo decidir si se trata de una visión trágica o cómica: el mundo determinista ha evolucionado a lo largo de millones y millones de años, produciendo eventualmente criaturas que crecen gradualmente en curiosidad y raciocinio hasta un punto fatal donde pueden ser causalmente llevadas a descubrir, en virtud de su misma racionalidad, lo inútil de sus proyectos y de sus frenéticas maquinaciones. Y así ingresan, en el espasmo final de un raciocinio que se aniquila a sí mismo, en un estado de absoluta estolidez. ¡Quizás esto mismo les sucedió a los árboles! Quizás, en los viejos tiempos los árboles deambulaban por el mundo, inmersos en sus proyectos, ¡hasta el terrible día en que se hizo la luz y tuvieron que echar raíces y "vegetar".[1]

Suponer que el veredicto de futilidad debe seguir inmediatamente a la confirmación del determinismo es confundir el tercer paso con el segundo. Este último (la afirmación de que el determinismo implica la impotencia causal de la deliberación) era claramente erróneo —en efecto, confundía determinismo con fatalismo—. El fatalismo es una concepción más bien mística y supersticiosa que afirma que, en ciertas encrucijadas de nuestra vida, nos hallaremos necesariamente en circunstancias específicas (las que "el hado" decrete), *independientemente de los caprichos o extravíos de nuestra trayectoria personal.* (Recordemos, por ejemplo, a los padres de Edipo, cuyo empeño en eludir la profecía creó las circunstancias para que ésta se cumpliera, "sellando" así sus destinos.) Todos coinciden en que esta clase de fatalismo no tiene nada de recomendable, como no sea crear efectos escalofriantes en la literatura. Determinismo no implica fatalismo y no hay razones para suponer que este último es verdadero.

Sin embargo, existen ejemplos genuinos de lo que podríamos denominar *fatalismo local*, a partir de los cuales se deriva la idea de que la deliberación es tan inútil como los movimientos del golfista una vez que la pelota ha iniciado su trayectoria. Pensemos en el hombre que se arroja del puente Golden Gate y, mientras va cayen-

[1] Según van Inwagen (1983, pag. 117), si uno no deliberara "echaría a correr o andaría a los tumbos sin concierto, o bien caería en un estado catatónico".

do, se pregunta "si habrá sido una buena idea". Ciertamente, las deliberaciones ya no le sirven en tales circunstancias. Podemos anticipar su futuro sin tener en cuenta las consideraciones que intervinieron en la resolución del problema. Sean las que fueren, no nos suministrarán por cierto una cadena causal que lo aparte de la trayectoria que hemos anticipado. Pero precisamente por tener esta característica, este caso es diferente de los casos habituales de deliberación.

En el generador de intuición que denominamos "el carcelero invisible", se encuentra una variedad de fatalismo local más afín a nuestras circunstancias normales:

> El hombre que está encerrado en un cuarto y no lo sabe puede elegir, desde luego, la permanencia, a pesar de no tener elección posible en ese aspecto. (Van Inwagen 1983, pág. 239.)

¿Por qué el carcelero invisible es un ejemplo más popular que el del puente Golden Gate? Porque no estamos seguros de no hallarnos en una situación semejante. No es habitual que sucedan estas cosas; pero, cuando suceden, podemos anticipar la trayectoria del habitante del cuarto sin tener en cuenta sus deliberaciones, tan impotentes como el engranaje que gira solo, desvinculado del resto de los engranajes. Cuando uno se da cuenta de que las propias deliberaciones están destinadas a desvincularse así de los senderos causales críticos, descubre verdaderas razones para la resignación, la apatía o la desesperanza. Pero no es lo mismo descubrir que se está atrapado en un torbellino de fatalismo local que descubrir que se vive inmerso en una corriente determinista.

Veamos un ejemplo de otra variedad de fatalismo local: el hombre que está loca y perdidamente enamorado y, lápiz y papel en mano, comienza a confeccionar una lista de los defectos de su amada, a fin de decidir si le propondrá o no matrimonio. (Charles Darwin lo intentó una vez.) Sabemos que ese hombre está perdiendo el tiempo; conocemos desde ya el resultado de sus patéticas deliberaciones; y no porque el resultado esté predeterminado, sino por el modo de determinarlo: demasiado simple e insensible a la información. Así como ignoramos el proceso mental del hombre que se arroja del puente cuando anticipamos su trayectoria, del mismo modo podemos ignorar los contenidos particulares del proceso mental del enamorado cuando anticipamos cuál será el veredicto, puesto que está funcionando un mecanismo compensatorio que transformará en virtudes los supuestos defectos de la amada: ella no es estúpida, tiene los pies en la tierra; ella no es egoísta, es vivaz. ¿Su rostro es poco agraciado? A algunas personas les gusta Picasso y a otras no. Nuestro enamorado llegará a un veredicto favorable con-

tra viento y marea.[2] Pero también en este caso, lo que lo diferencia de los casos comunes de deliberación es, precisamente, el elemento que justifica la interpretación fatalista: la sordera del deliberador a la voz de la razón (recuérdese que está "loca y perdidamente" enamorado).

La capacidad del enamorado para deliberar engañándose a sí mismo es, ciertamente, un caso extremo y todos sucumbimos a esta tentación, con mayor o menor frecuencia y en diferentes grados. Esta patología universal, leve o severa, se manifiesta de distintas formas y cubre un amplio espectro que va desde el polo del *autoengaño* hasta el de la *falta de voluntad* (*akrasia*, según Aristóteles). Considérese al fumador: en la medida en que no ponga fin a los retorcidos y astutos argumentos que utiliza para convencerse a sí mismo de que, a pesar de todo, fumar es un pasatiempo placentero, elegante, que calma los nervios, favorece la concentración mental y tiene escaso margen de riesgo, su patología reside en el autoengaño. En la medida en que deseche tales sofisterías y se persuada de la flagrante estupidez de fumar —y continúe fumando—, su patología reside en la falta de voluntad. Cuando existe una tensión entre la inclinación y el interés manifiesto de este tipo, se produce una suerte de transacción entre la totalidad de las operaciones de la maquinaria pensante y el grado en que participan los engranajes que controlan la conducta. Las personas "conducen" de distintas maneras en circunstancias distintas: a veces, "aprietan el embrague" y entonces razonan bella pero ineficazmente; a veces "funden el motor", las deliberaciones se degradan y el pensamiento se desintegra bajo el peso del deseo.

La metáfora no hace justicia al carácter confuso y complejo del fenómeno, que ha sido investigado intensamente en filosofía (véanse, por ejemplo, Davidson 1970, de Sousa 1970, A. Rorty 1972, Schiffer 1976). Hay escaso consenso con respecto a cuál sería el análisis correcto de dicho fenómeno, pero todos coinciden en un punto, que es lo único que nos interesa aquí: aunque nadie sea inmune a esta condición, se consideran anormales los casos severos y verdaderamente inhabilitantes, como la aflicción del enamorado o, en el otro extremo, la compulsión del adicto, tan poderosa que lo pone fuera de control. Pero la condición humana no está *amenazada* por este tipo de fatalismo local.

[2] El conjunto relevante de contrafácticos que consideraríamos verdaderos respecto de él tiene un consecuente común; esto es, suponemos que, salvo excepciones poco significativas, obtenemos enunciados verdaderos sustituyendo a *p* por diferentes proposiciones en "si se le señalara que *p*, decidiría igualmente proponerle matrimonio a Lulú".

Sabemos, pues, que hay situaciones en las que toda deliberación es inútil, pero también sabemos que esas situaciones son, por fortuna, anormales en nuestro mundo. Así, al dejar atrás la doctrina del fatalismo global o universal (nuestro segundo paso), también debemos abandonar la moraleja que es válida sólo si (o cuando) el determinismo es verdadero. Las deliberaciones *son* (en general) efectivas en un mundo determinista pero no fatalista. ¿Cómo podríamos concluir racionalmente que es inútil deliberar en una circunstancia en la cual nos asisten todas las razones para creer que si el resultado de nuestra deliberación es la decisión de hacer A, entonces haremos A, y si el resultado de nuestra deliberación es la decisión de hacer B, entonces haremos B y, además, es muy probable que el resultado de nuestra deliberación sea el único acto racional que cabe, teniendo en cuenta las circunstancias?

Sin embargo, aún persiste el sentimiento de que es absurdo seguir deliberando una vez que se tiene la creencia de que el deliberar mismo está tanto en el proceso como en el producto final.[3] Tenemos el sentimiento de que si alguien continúa deliberando en tales circunstancias, *algo* está engañándolo. Por lo tanto, ante la evidencia, o bien tendrá que revelar un punto ciego semejante al de la *sphex* —es decir, un rechazo o una incapacidad para aceptar la verdad devastadora—, o bien tendrá que hacer lo más racional y honorable: hacerse un *hara kiri* intelectual. Pero supongamos que la persona continúa deliberando alegremente y se desembaraza de la acusación de irracionalidad mediante la conocida réplica: "Si soy tan estúpido, ¿por qué tengo tanto éxito?". Es, sin duda, una réplica tentadora, pero es posible que sólo se trate de una expresión de deseo. ¿Cómo podríamos saberlo?

Recordemos al golfista del capítulo 1, que mantenía la cabeza baja durante el *follow-through*, después de que la pelota abandonó la cara del palo. Se puede aducir que esta política es irracional y supersticiosa, ya que la trayectoria de la pelota ya está determinada antes de que comience el *follow-through*. No obstante, si reflexionamos, veremos que esa política posee un sólido fundamento racional. Mirar hacia adelante con la cabeza baja y no ceder a las distracciones tiende a asegurarnos que sucederá lo correcto en el momento crucial del *swing*. ¿Podría existir un fundamento racional

[3] La exageración del primer paso se repite a menudo aquí; se declara imposible lo que *podría* ser absurdo o irracional que suceda: "Si uno ignora lo que hará pero sabe que ya existen condiciones suficientes para hacerlo, entonces no podrá deliberar acerca de lo que habrá de hacer aun cuando no sea posible saber cuáles son esas condiciones" (Taylor 1964). Esto es obviamente falso, a menos que se interprete "deliberar" de tal manera que nada puede ser considerado como una deliberación irracional o ilusoria, en cuyo caso no sería tan obviamente falso.

parecido en favor de la política de mirar hacia adelante de una cierta manera y de no ceder a las distracciones cuando la deliberación está en pleno movimiento?

2. *El diseño del deliberador perfecto*

> *La vida sólo puede ser comprendida hacia atrás;*
> *pero debe ser vivida hacia adelante.*
> —Kierkegaard
>
> *Si bien es parte de nuestra naturaleza*
> *producir futuro, el sistema está ensamblado*
> *de tal manera,*
> *que nuestras predicciones han de permanecer*
> *dudosas. No podemos pensarnos sin el instante*
> *siguiente, pero ignoramos*
> *cómo será ese instante.*
> —François Jacob (1982, pág. 67)

Al parecer, es inevitable que actuemos movidos por la idea de libertad. Nos empeñamos en deliberar como si tuviéramos un futuro abierto. Pero ¿hay algo censurable o lamentable en esa tendencia irresistible? Adoptando el punto de vista de la Madre Naturaleza, podemos evaluar la racionalidad o la prudencia que se ocultan en nuestras deliberaciones, o bien preguntarnos acerca de las delimitaciones de un deliberador físico finito, con el fin de probar la hipótesis de que el determinismo es compatible con la deliberación (racional). Vamos a suponer entonces que el determinismo es verdadero y veremos si es necesario negarlo en algún punto ulterior para que tenga sentido nuestra empresa. ¿Qué características le asignaríamos al deliberador perfecto, en caso de tener que diseñarlo?

Como vimos en el capítulo anterior, debe tener la capacidad de ser movido por los elementos actuales de su entorno, de modo de prever el medio ambiente causal al que ingresará muy pronto; y debe preverlo con rapidez, seguridad y precisión. Existe, sin embargo, una especie de transacción entre estos *desiderata*. Si ha de ser una auténtica previsión y no una percepción tardía, habrá que pagar un precio: el dispositivo simplificará las tareas relativas al manejo de información.[4] Debe desechar e ignorar los datos que pue-

[4] El precio no es, en realidad, mucho más alto. Como señaló Simon (1969, pág. 110) en un brillante debate filosófico sobre estos problemas: "Una descripción tolerable de la realidad sólo necesita tener en cuenta una pequeña fracción de todas la posibles interacciones".

de obtener en principio, pues los costes (en tiempo y esfuerzo) de acarrear demasiada información acaso anulen el plazo requerido para que la previsión se realice.[5] Y no sólo debe desechar algunos datos: si debe mantener un *modicum* de predicción confiable, habrá de efectuar las simplificaciones correctas, descartando los datos "correctos". En esta transacción, el equilibrio de los factores constituye una materia de índole peculiar que varía de una especie a otra, dependiendo de sus circunstancias o talentos. El oso hormiguero, a diferencia del pájaro insectívoro, no invierte sus energías en rastrear insectos individuales, probando su trayectoria y prediciendo con sutileza su futura posición, sino que se limita a calcular al tanteo el área infestada de hormigas y a dejar que su lengua se apodere de las más lentas.[6] Un filósofo podrá aducir que "hormiga" es un *término de masa* en el lenguaje mental del oso hormiguero y que algunas zonas contienen "más hormiga" que otras.[7]

Toda transacción específica entre estos factores dictada por las circunstancias de una especie, comportará ciertas distinciones. En cada deliberador, pues, se trazará una línea anulando los elementos del entorno que "considere":

(1) *fijos,* tan absolutamente fijos que no necesite desperdiciar esfuerzos en seguirles la pista;

(2) *indignos de atención,* tan irrelevantes para los intereses del sistema que cabe ignorar si son o no fijos sin correr ningún riesgo.

(3) *cambiantes (y dignos de atención).*

Los elementos cambiantes deben diferenciarse, además, según esos cambios sean

(3a) *manejables, al menos en ciertas condiciones;* y por lo tanto predecibles provechosamente en dichas condiciones; o

(3b) *caóticos,* caprichosos o "fortuitos", impredecibles pero relevantes y dignos de atención. (En este caso, "impredecible" signi-

[5] "Predecir el futuro de una curva es efectuar una cierta operación sobre su pasado. Ningún aparato construible puede ser una realización del operador de predicción verdadera; pero hay ciertas operaciones similares que sí pueden ser realizadas por un aparato construible" (Wiener 1948, pág. 12).

[6] El ejemplo, como muchos de las ideas que aparecen en esta sección, pertenece a Wimsatt (1980); véase especialmente la pág. 296.

[7] A grandes rasgos, los términos de masa se contraponen a los términos de número o de clase: se pregunta *cuántas* sillas hay en el cuarto, cuántas estrellas en el cielo, cuántos minutos en una hora; en cambio, se pregunta *cuánta* azúcar hay en el recipiente, cuánta agua bajo el puente o cuánto tiempo falta para la cena —algunos términos se pueden usar de ambas maneras, según el contexto—. El ejemplo de Quine (1960): Mary tenía un corderito [un poco de salsa y un puré de patatas...]. Para problemas y detalles más específicos sobre este punto, véase Cartwright (1970).

fica prácticamente impredecible: una vía de recursos de la que se puede prescindir.) [8]

Los acontecimientos cambiantes dignos de atención (predecibles o impredecibles) deben ser puntualmente localizados. No se los puede calcular al tanteo como hace el oso hormiguero con las hormigas, ya que sus trayectorias reales y específicas tienen mucha incidencia en el cálculo del bienestar del sistema. Pero no siempre se les puede seguir la pista y algunos de ellos no son rastreables en absoluto.

Esta incertidumbre tiene que controlarse; a pesar de los conocimientos del que predice, se debe desmembrar el área de las coordenadas espacio-temporales que ocupan las cosas no rastreables y distinguir, de la base de las coordenadas excluidas, los estados "imposibles". Por otra parte, es necesario, para ahorrar esfuerzos cognitivos, dividir *grosso modo* este conjunto de alternativas en un número pequeño y manejable de diversos *estados* posibles, calculando al tanteo, como se calculan las hormigas desechables del enjambre, las microdiferencias dentro de cada estado. Según Winsatt, "en el contexto de tales modelos ni siquiera un organismo omnisciente y omnipotente sería un demonio laplaceano. No necesitaría buscar el conocimiento total del universo sino el de aquellos factores que afectan a la adaptación" (1980, pág. 239).[9]

La variedad de posibilidades que así se introduce en el esquema conceptual de todo deliberador bien diseñado es, por cierto, la variedad de las *posibilidades epistémicas*. Si bien individualiza los estados de cosas que los deliberadores pueden conocer y de las cuales pueden ocuparse, da origen a un concepto cuyo pasado epistémico habitualmente se soslaya: el concepto de las *potencialidades* de una cosa.[10] Los perros *pueden ladrar,* pero dónde y cómo lo hacen es impredecible las más de las veces. Para algunos

[8] Véase la exploración precursora de esas ideas en Wiener (1948, págs. 43 y siguientes), sobre el "tiempo newtoniano y bergsoniano". Considérese el "supuesto de rigidez" aparentemente enclavado en las operaciones de nuestro sistema visual, como un ejemplo de un supuesto implícito respecto de un elemento *fijo* (Ullman 1979). El término "caprichoso" se debe a Richard Lewontin. Véase Wimsatt 1980, pág. 289.

[9] Wimsatt es imprudente al referirse a un organismo "omnisciente y omnipotente", pues ese ser infinito puede ignorar las limitaciones que abundan en *cualquier* fundamento racional. Pero se puede interpretar lo que quiere decir: un organismo finito, bien diseñado, no encontraría que sus circunstancias mejoran en lo más mínimo por el hecho de concederle poderes laplaceanos. Lo que está en juego es el valor y la necesidad práctica de adoptar la actitud intencional con respecto a uno mismo y a los demás agentes (Dennett 1981b). Para otras posibles alternativas y variaciones sobre este punto, véanse Simon (1960) y Hofstadter (1979).

deliberadores, que un perro del vecindario ladre en el tiempo *t* es un hecho tan digno de atención como indeterminable. Ryle afirma que "decir que un perro ladra significa que uno no puede *fiarse* de su silencio". Estos mismos deliberadores que tanta importancia dan a los ladridos no se ocupan en cambio de cómo ladra un perro. Piensan que cuando se lo ha escuchado a uno, se los ha escuchado a todos. Por consiguiente, consideran a los perros como sistemas cuyo grado de libertad posee dos estados y los dividen en dos categorías: la de los ladradores "encendidos" y la de los ladradores "apagados". Este concepto simplista de los perros —simplista en extremo para algunos propósitos pero no para todos— les permite plantear un problema potencialmente resoluble en la teoría del control: ¿puedo inducir al ladrador al estado "apagado" y mantenerlo en él indefinidamente?

Un *tipo* peculiar de agente deliberador —digamos una especie— siempre estará equipado con una manera propia y específica de recolectar y dividir la información del entorno; tendrá, pues, una manera de "concebir" el mundo que le permitirá actuar con eficiencia en su mundo. Ampliando un poco más el concepto de Sellars, denominaremos a ese "esquema conceptual" de las especies su *imagen manifiesta* (¿visible?) (Sellars 1963, véanse especialmente las págs. 6-14). Algunos elementos de la imagen manifiesta son condiciones básicas, permanentes, en tanto que otros son "posibles" estados de cosas en el mundo. Entre los posibles estados de cosas, los hay impredecibles, seguramente predecibles (por diversos agentes deliberadores) indirectamente controlables por (los efectos de) las acciones de los deliberadores, y algunos —las acciones mismas [11]— directamente controlables por el deliberador.

El manejo prudente de la información permite al deliberador tomar decisiones útiles y seguras sobre el control antes de que sea demasiado tarde. Tales decisiones no sólo deben basarse en los objetivos (o deseos) del deliberador, sino también en expectativas fiables de diversa índole:

[10] Véase Hofstadter (1979) y, también, su discusión sobre la esfera contrafáctica implícita y los "ejes de deslizamiento" en Hofstadter (1982c).

[11] Sería más acertado llamarlas "acciones básicas" del deliberador (siguiendo a Danto 1965), pero hay que tener en cuenta que el conjunto de acciones que son básicas para un agente o un deliberador puede cambiar como resultado del entrenamiento o de otra experiencia. Hay que tener cuidado de evitar todo tipo de fundamentalismo basado en las acciones "básicas", pues, si bien es cierto que en cualquier instante es posible distinguir entre aquello que el agente puede decidir hacer "directamente" y aquello que el agente puede decidir (intentar) producir haciendo "directamente" otra cosa, la pertenencia, fácilmente variable, a estas clases hacen que resulten cimientos precarios para una teoría fundamentalista.

(1) lo que sucederá pronto, a pesar de lo que haga el deliberador;
(2) lo que sucederá pronto si el deliberador hace *A* o *a menos que* el deliberador haga *B*.

Si el objetivo específico del deliberador es la preservación y protección de *x*, entonces, las circunstancias probables y los estados futuros de *x* ocuparán un lugar prioritario en sus cálculos, especialmente si *x* es el deliberador mismo. De esa manera, la autopredicción es un componente esencial en la subsistencia de un deliberador que se autocontrola, como hemos visto, por otras razones, en el capítulo anterior. Un deliberador de esta índole debe tener una idea clara respecto de aquello que prevé. Pero la auto- predicción se divide en multitud de categorías y es limitada. Existen expectativas acerca de:

(1) lo que le ocurrirá pronto al deliberador a pesar de lo que haga;
(2) lo que le ocurrirá pronto al deliberador si hace *A* a menos que éste haga *B*;
(3) lo que el deliberador pronto decidirá hacer.

Tener expectativas acerca de las dos primeras clases no es más difícil —tampoco más fácil— que tener expectativas paralelas acerca de otras cosas. Pero las expectativas en cuanto a lo que el deliberador decidirá pronto son más problemáticas (Ryle 1949, pág. 188; Popper 1951; MacKay 1960; Pears 1964). Cualquier intento de rastrear el propio proceso de deliberación para efectuar una proyección precisa de su trayectoria se anulará a sí mismo, amenazando con una regresión infinita de autoanálisis. Pero esto no significa que no haya ningún progreso en cuanto a estrechar la esfera de las propias decisiones (Pears 1964). El deliberador puede prever —exceptuando la interferencia imprevista— que decidirá seguir uno o varios cursos de acción (individuados a *grosso modo* e ignorando las microvariaciones dentro de cada caso). Cualquiera de estos cursos emergerá realmente, mientras el curso adoptado permanece imprevisible para el deliberador, quien los considera a todos como "posibilidades".

En consecuencia, deben existir dos importantes categorías de acontecimientos que sean siempre imprevisibles para el deliberador: los acontecimientos que atañen a cosas de su mundo que son impredecibles en la práctica (pues las restricciones superan el "gasto") y los acontecimientos que atañen a sus propias deliberaciones y que trascienden su maquinaria de prognosis de un modo tal que ninguna inversión extra en maquinaria cognitiva podría superarla.

Hay categorías inevitables de acontecimientos futuros que son epistémicamente posibles o epistémicamente abiertas al agente. Nuestras propias decisiones no son sistemáticamente impredecibles, aunque otras personas las puedan anticipar.[12] Obtendremos, forzosamente, un rédito menguante de cualquier gasto creciente que insuma el rastreo o cálculo en esta área de predicción. La sabia moraleja que todo deliberador o deliberador-diseñador debería extraer es: "ni siquiera intentarlo", pues se trata de una práctica que se anula a sí misma. La única oportunidad de llegar a la decisión deseada es "mantener la cabeza baja" (como el golfista) y continuar deliberando. Se trata de un buen consejo, si se desea cosechar los beneficios de la deliberación. Algunos agentes "se despreocupan", sin embargo. (Pero si alguien prefiere jugar a Tristram Shandy con sus propios estados mentales, precipitándose en una espiral infinitamente regresiva de introspección y permitiendo que el mundo le aseste todos los golpes que pueda, ése es otro asunto.)

Por lo tanto, la imagen manifiesta de todo deliberador incluirá la división (que aparece como resultado de sus deliberaciones) en cosas que son de su "incumbencia" y en cosas, previsibles o imprevisibles, fijas o mudables, que no son de su incumbencia. Esta apertura epistémica, esta posibilidad "según lo que se sabe", nos brinda la libertad de acción que requieren nuestras deliberaciones. Nos encontramos, pues, en una posición en la que existe un sólido fundamento racional en pro de esta política: aunque la propia decisión esté determinada, ignoramos cuál será esa decisión.

Este elemento capital de la visión del mundo o de la actitud del liberador, a menudo advertido pero rara vez analizado por los filósofos, ha comenzado a recibir otro tipo de análisis teórico por parte de los científicos de la inteligencia artificial. McDermott (1982), en su interesante artículo "A Temporal Logic for Reasoning about Processes and Plans", ejemplifica un simple problema de planeamiento presentado en "primera persona", pero pensado como un ejercicio para colaborar en el diseño de los robots que solucionan tales problemas. (En los párrafos siguientes, mis opiniones se mezclan con las de McDermott.)

Supongamos que quiero pintar el techo de mi cuarto y la escalera de mano. Cuál de los dos debo pintar primero es, ciertamente, un asunto de mi incumbencia. Pero abrir o no la lata de

[12] ¿Qué ocurre si otra persona trata de decirme su predicción antes de que yo haya tomado una decisión? O bien confundirá su predicción con respecto a mí con su propia incapacidad de autopredicción (qué le diré, cuándo y cómo), o bien (si intenta manejar este problema "dirigiendo su blanco") fabricará una predicción cuya garantía depende de que la crea cuando me la dice. En ambos casos, la incertidumbre sobre mi propio futuro es ineliminable. Véanse las discusiones en Popper 1951 y MacKay 1960.

pintura antes de mojar el pincel o usar o no la escalera para pintar el techo, son cosas que no me incumben. Por supuesto, *podría* imaginar otras formas de llegar hasta el techo o de extraer pintura de la lata, pero no lo hago y eso es lo importante. Ni siquiera trato de imaginar las "posibles maneras" de resolver el problema (o de reemplazarlo por otro). Considero que ciertas condiciones están dadas, que no son de mi incumbencia en esta ocasión... y continúo con mis planes.

Tampoco me incumbe si, cuando pinto el techo corro el riesgo de que la pintura chorree por las paredes (modificar la dirección de la gravedad no fue jamás un asunto de mi incumbencia). Ahora bien, si estoy pintando una silla me importará su posición ya que puedo ponerla al revés, precisamente para que la gravedad me ayude a evitar las chorreaduras. También *podría* arreglármelas para poner el cuarto al revés antes de pintar el techo. Es una manera (formalmente posible) de hacerlo, pero ningún yo podría o debería considerar seriamente la posibilidad; lo mejor es ignorar el asunto por imposible, como si perteneciera a la categoría de las cosas inmutables que no merecen nuestra atención.[13]

Lograr una esfera de acción óptima en el área de nuestra incumbencia y un tipo de posibilidades ni demasiado amplio ni demasiado restringido que nos permita invertir algún precioso esfuerzo de deliberación, constituye un problema crítico y nada trivial para cualquier deliberador (o deliberador-diseñador). Si, por un lado, es correcto no pensar en dar vuelta el cuarto antes de pintar el techo, por el otro lado, quedaríamos anclados en la rutina sin esos saltos de la imaginación. Sartre advirtió su importancia y con su acostumbrada y fría reticencia definió al agente libre como "un ser que puede producir una ruptura anonadante con el mundo y consigo mismo" (Sartre 1956, pág. 435). "El dio en que podemos concebir un estado de cosas diferente, una nueva luz alumbra nuestros problemas y nuestros sufrimientos y *decidimos* que son insoportables." (Sartre 1956, pág. 439.)

Una criatura con una política obstinada y poco intrépida del manejo de la imaginación en su imagen manifiesta no correrá jamás el riesgo de enfrentarse al problema del libre albedrío (y menos aun el de ser perturbado por él). Pero los seres humanos han probado ser muy listos. Su imagen manifiesta es el mundo cotidiano de las cosas coloreadas, no el enjambre de moléculas. Nuestros ojos son macroscopios, no microscopios. Percibimos los cambios de "tamaño mediano"; las cosas que ocurren muy rápida o muy lentamente resultan imperceptibles para nuestros ojos y oídos. Pero a diferen-

[13] Véase Levi 1980, sobre el concepto de *posibilidad seria* de un agente.

cia de otras especies, poseemos, además, lo que Sellars denomina "la imagen científica". Quizá no estemos naturalmente bien equipados para *ver* o *seguirles la pista* a los electrones, pero somos capaces de comprender que si el "agua" es un sustantivo de masa, también es un enjambre de moléculas que pueden ser contadas y cuyas trayectorias son manejables en principio y a veces en la práctica (con la ayuda de las prótesis que amplian nuestros sentidos).

Dada la vastedad de nuestras circunstancias y propósitos, nos favorece incluso abandonar la prudente economía de nuestra imagen manifiesta y volver a concebir parte de nuestro mundo en un plano descriptivo más sutil y diverso. Es entonces cuando descubrimos lo incompleto del esquema conceptual de nuestra imagen manifiesta y comenzamos a considerarlo una perspectiva limitada, prejuiciosa y revocable del mundo. Vimos que nuestra capacidad para participar en deliberaciones de tiempo limitado —aun las requeridas para comprometernos en la investigación científica— depende de nuestra imagen manifiesta, pero aún parecería ser una especie de ilusión nacida de nuestra pobreza cognitiva.

Nos encontramos, pues, en un dilema aparente, fluctuando entre una manera práctica de pensar acerca del mundo —acaso práctica en grado óptimo— y una visión heredada que si bien no es práctica, todavía es racional. Hemos progresado, sin embargo, al encontrar razones que sustentan el presentimiento de que nuestra *manera de pensar* respecto del sitio que ocupamos en el universo no es sólo la única manera posible (sobre la base de lo cotidiano) sino la única que puede pensar un deliberador racional, finito, sea o no verdadero el determinismo.

Si queremos deliberar bien, es racional que obremos *como si* el mundo tuviese un futuro abierto, con posibilidades reales.

Pero, ¿ello no equivale a decir que *es racional mantener la ilusión* de un futuro abierto, con posibilidades reales? No necesariamente. Yo supuse, en favor del argumento, que estábamos discutiendo sobre deliberadores deterministas en un mundo determinista, pero no supuse, en favor del argumento, que esos deliberadores actuaran en un mundo carente de oportunidades reales. No obstante, observando más de cerca el concepto de oportunidad, cabe llegar a la misma conclusión de otra manera.

3. Oportunidades reales

¿En qué consiste una oportunidad real? ¿Nuestras oportunidades serán sólo aparentes si el determinismo es verdadero, y reales si el determinismo es falso? Se supone que una oportunidad

es la "posibilidad" para un agente de "hacer algo" que establezca una diferencia". Las posibilidades deseadas son aquellas en las que el determinismo se suspende (si acaso rigiera) y cuya *posibilidad* es real, ¿o puede haber posibilidades en un mundo perfectamente determinista, con deliberadores deterministas (o "deliberadores")?

Vamos a partir de un mundo determinista para observar qué sentido tienen las oportunidades en ese caso. Nos ocuparemos del mundo del robot explorador con el objeto de saber exactamente lo que estipulamos al afirmar que su sistema de control es determinista; lo hemos diseñado con ese propósito y, además, para que sea muy resistente a los ruidos de micronivel y a las perturbaciones fortuitas. Su estructura no sólo carece de contadores Geiger que generen efectos aleatorios sino que fue diseñado para amortiguar tales efectos. Está girando, por otra parte, en un planeta que se supone deshabitado; es decir, incontaminado por agentes auténticos que, de ser macro-indeterministas, podrían introducir un indeterminismo en gran escala en el entorno donde opera el robot. (Otro motivo que justifica la elección del ejemplo: como no nos importa que el robot tenga oportunidades reales, es improbable que nos embarquemos en razonamientos que no son sino mera expresión de deseos.)

Supongamos, entonces, que el Deliberador Determinista Mark I está diseñado para explotar al máximo sus oportunidades. ¿Cuáles son esas oportunidades? El robot tiene intereses (planeados, artificiales) en el sentido analizado en el capítulo dos. Los acontecimientos relativos a esos intereses le interesan específicamente; en consecuencia:

> El robot tiene una oportunidad de robot cuando esos acontecimientos que específicamente le interesan lo llevan a "evaluar", y si así lo decide, a planear y ejecutar, en un tiempo limitado, un proyecto diseñado por él mismo con el fin de ampliar sus intereses a la luz de esos acontecimientos.

En nuestro mundo cotidiano, sea o no determinista, existen dos maneras de negarle oportunidades a la gente. La primera y más efectiva es la fuerza bruta: se erigen barreras o restricciones de manera que resulte imposible hacer algo, *aunque se lo desee*. En ocasiones, el método se utiliza para evitar daños. Por ejemplo, en las fábricas donde los obreros deben colocar piezas en las fauces de enormes máquinas para que éstas les den forma, se les exige el uso de una especie de puños atados por fuertes cadenas, que a su vez se atan a una palanca que es parte de la máquina. A medida que las cuchillas descienden, levantan automáticamente la palanca, cuyas cadenas son lo bastante cortas para sacar por la fuerza las manos del obrero, si éste no las ha sacado "por propia y libre voluntad",

evitando así el riesgo de una mutilación. Incluso si el obrero decide (por extraños motivos) poner las manos dentro de la máquina en el momento equivocado, no podrá hacerlo.

Este tipo de restricciones impuesta por la fuerza bruta anonada las deliberaciones, creando islas de fatalismo local en donde virtualmente se garantiza la ineficacia de la deliberación. Del mismo modo, la prisión mantiene a los presos en un estado en el cual no caben las decisiones, salvo las más triviales: dónde se sentará en la celda, etcétera.

Pero imaginemos un carcelero excéntrico que una vez al mes y por la noche abre las cerraduras de todas las celdas y la del portón de la cárcel, luego de comprobar que los presos están profundamente dormidos. Al parecer, el carcelero ha creado deliberadamente una oportunidad de fuga, y así lo considera el consejo de alcaldes de la prisión cuando lo releva del cargo. Pero el carcelero aducirá que los presos nunca tuvieron *realmente* la oportunidad de escapar, ya que siempre esperó a que estuvieran dormidos y nunca les informó sobre la existencia de la "oportunidad". Y ésta es la manera de negarle oportunidad a la gente: mantenerla en la ignorancia de que existe una oportunidad. Denominaremos *oportunidad rasa* (¿desnuda?, ¿lisa y llana?) a este tipo de oportunidad no reconocido ni imaginado. Si camino por una calle en donde hay una hilera de tachos de basura y uno de ellos contiene una bolsa de diamantes, habré perdido una oportunidad rasa de volverme rico. El hecho de que no tenga razones para suponer la existencia de piedras preciosas al alcance de la mano, o que mi comportamiento normal no incluya el revolver tachos de basura en busca de objetos valiosos, no cambia las cosas en absoluto.

Las oportunidades rasas abundan pero no son suficientes; cuando hablamos de oportunidades o posibilidades de mejorar nuestra vida no nos referimos, por cierto, a las oportunidades rasas. Queremos detectar nuestras oportunidades o que se nos informe con tiempo de su existencia, de modo de poder aprovecharlas.[14] Por ese motivo, se objeta con justicia toda asignación de becas por concurso abierto que no se publicite adecuadamente: sólo los que estén enterados y se presenten antes de que expire el plazo, tendrán una oportunidad real de ganar el concurso.

Existe una gran similitud entre el caso de la máquina cortadora, diseñada para *eliminar* sistemáticamente una oportunidad es-

[14] "Si nos remontamos a una época y a una cultura muy distantes de la nuestra, admitiremos que las posibilidades de acción, basadas en discriminaciones no reconocidas entonces, no eran genuinas posibilidades de acción para quienes vivían en aquellos tiempos" (Hampshire 1959, pág. 184).

pecífica, y el del carcelero, diabólicamente sistemático en *conceder* una oportunidad. El obrero de la fábrica, en el transcurso regular de su tarea, siente con frecuencia el deseo de introducir las manos en la máquina para colocar o retirar una pieza. Pero hay ocasiones, afortunadamente predecibles, en las que resultaría seriamente dañado si actuara según su deseo. Por esta razón, la máquina fue diseñada para frustrarlo toda vez que aparece en la circunstancia equivocada. Se supone, por otro lado, que los presos experimentan un deseo constante de escapar y, en este caso, el carcelero les brinda la oportunidad pero sólo cuando no es posible identificarla y, por lo tanto, aprovecharla.

Ambos métodos son "perfectibles". La máquina cortadora puede tener un monitor cerebral que registre el estado actual del deseo del obrero y sólo tire de la cadena cuando detecta un deseo surgido a destiempo. Del mismo modo, cabe que el carcelero eche mano de un sistema más seguro para ventilar a los presos, conectando las cerraduras de tal manera que sean directa y electrónicamente controlables por los EEG y los REM del sueño de los presos; al menor signo de despertar, las cerraduras volverán a cerrarse.[15] Lo importante, en tales casos, es que el agente que controla alguna faceta del mundo de otro, lo hace vinculando sistemáticamente las condiciones; por lo tanto, nunca se satisfacen al mismo tiempo todas las condiciones necesarias para actuar según la oportunidad (mejor, para aprovecharla). A diferencia de estas tortuosas criaturas, el entorno no oculta sistemáticamente las oportunidades. La naturaleza no juega a las escondidas: es sutil, pero no maliciosa, como dijo Einstein respecto de Dios.

Una oportunidad real es, pues, una ocasión en la cual el agente que se autocontrola "enfrenta" (o es informado sobre) una situación en donde las "deliberaciones" ulteriores constituirán el factor decisivo. En tales situaciones, cabe más de una alternativa para el agente o el que se autocontrola; es decir, el nexo crítico pasa por la

[15] Esta ciencia ficción podría fácilmente llevarse a la práctica. Un anticipador real aún más extraño ya está en funcionamiento en la Universidad de Illinois. Lo llamo la máquina de Berkeley (Dennett 1982b). Un sujeto se sienta frente a la pantalla de un ordenador y lee el texto que allí aparece. Un dispositivo óptico extraordinariamente preciso registra permanentemente los rápidos movimientos de los ojos —o *saccades* — del lector. Una *saccade* es un movimiento "balístico" y su trayectoria puede trazarse poco después del lanzamiento lo mismo que la de un cohete. La máquina hace eso y, en un tiempo determinado, el ordenador calcula en qué lugar de la página se posarán los ojos y, con más rapidez que un mago, borra la palabra y la reemplaza por otra. El lector no advierte ningún cambio. Para el lector, la página podría estar grabada en mármol. Pero otro lector, que mira la pantalla por encima de su hombro, moviendo los ojos a un ritmo diferente, ve la página algo movida a causa de los cambios de palabra.

deliberación. Cuando se dice que el resultado es de la incumbencia del agente, debe entenderse en este sentido. (Una vez que el robot está fuera de nuestro alcance, girando en torno del planeta deshabitado, queda librado a sí mismo y lo que haga de ahí en más para resolver los problemas que se le presenten, será de su exclusiva incumbencia.)

La descripción del robot incluye la facultad de aprovechar sus ocasiones y actuar en la mayor parte de las oportunidades. Lo hará mientras ningún robot malicioso, capaz de predicción, comience a disponer las cosas en fase con las pautas exploratorias. Pero ¿qué ocurre cuando el robot está en condiciones de actuar correctamente de acuerdo con la información recibida y simplemente no lo hace? Aunque el robot sea altamente fiable, y esto sucede en raras ocasiones, no es perfecto; si ha fracasado, es porque estaba determinado para fracasar, pues, *ex hypothesi*, el robot está siempre determinado. ¿Habrá tenido una oportunidad *real* en esa instancia? Si algunos de nuestros camaradas humanos fracasara en la ocasión, diríamos simplemente que la ha desperdiciado: tuvo la oportunidad y no supo capitalizarla. Cabe que no sea ésta una forma adecuada de expresarse respecto de los seres humanos, pero ¿no es obvio que el robot, en virtud de su naturaleza determinista, estaría incorrectamente descrito en esos términos?

El robot fracasó en la resolución del problema porque no atinó con el razonamiento justo o bien porque no le asignó la debida importancia a la información que poseía. El robot es finito e incapaz de seguir la política de evaluar permanentemente todo aquello que podría ser relevante para sus intereses. Pero en tanto que autocontrolados bien diseñado y de tiempo efectivo, utiliza procedimientos heurísticos, es decir, está planeado para "correr el albur", confiando en algún tipo de evaluaciones fortuitas. Para obtener este elemento de diseño en un sistema determinista, se lo equipa con una subrutina perfectamente determinista que genera, cuando se lo requiere, una secuencia de dígitos virtualmente sin pautas. Casi todos los ordenadores en la actualidad vienen equipados con alguna clase de "generador de número aleatorio", pero la secuencia generada por el dispositivo no es aleatoria sino pseudoaleatoria; o sea que no es el efecto de un proceso cuántico indeterminista y no es matemáticamente aleatoria en lo que se refiere a la incomprensibilidad de la información (Chaitin 1976).

Cuando analizamos microscópicamente el programa del robot en el momento de la falla, advertimos que si el programa que genera el número aleatorio dio un 3 y ya estaba determinado, por supuesto, que el número a generarse sería el 3. Luego, ¡el robot no tuvo jamás una oportunidad!

Puesto que usted se siente inclinado a suponerlo así, giraremos entonces la perilla de nuestro generador de intuiciones para intentar una variación o mutación: en el Deliberador Determinista Mark II, se reemplazará el programa que genera el número pseudoaleatorio por un auténtico generador de radio del número aleatorio (incorporando a la estructura un contador Geiger que registre la caída de un poco de radio), de ese modo, ni siquiera la misma inteligencia superior de Laplace, con su infinita capacidad de predicción, sabrá de antemano si el robot puede o no aprovechar la oportunidad. ¿Se trata, en verdad, de un procedimiento más justo? ¿Tendrá el robot una oportunidad real en esta instancia?

Si su intuición le dice que sí, entonces significa que hemos progresado mucho. No sólo cabe mejorar la categoría existencial del robot por medio de un programa que genere el número aleatorio equipado con un contador Geiger o con cualquier otro amplificador de efecto cuántico, sino afianzar nuestras esperanzas respecto del libre albedrío, sobre la base del descubrimiento de un *"hardware"* similar que forma parte orgánica de nuestro cerebro (Lucas, 1970). El Deliberador Mark II no supera, ni en el corto ni en el largo plazo, la capacidad de acción del Mark I; tampoco lo aventaja en lo que se refiere a la supervivencia, pues no existen malignos demonios laplaceanos en su entorno, pero usted piensa que, al menos, está mejor ubicado *metafísicamente*: a diferencia del Mark I, sus oportunidades serán reales.

Analicemos con más cuidado el trasfondo de tales intuiciones. ¿Por qué sería más real la "posibilidad" en el robot equipado con el contador Geiger que en el robot cuyo programa genera el número seudoaleatorio? La intuición que supone una diferencia de oportunidades entre ambos robots es ilusoria y sus falsas credenciales quedarán al descubierto apenas las consideremos en otro contexto.

Comparemos ahora la imparcialidad de dos clases de lotería. En la lotería A, el sorteo se lleva a cabo luego de la venta de todos los billetes. Se colocan entonces los respectivos talones en un mezclador, se los mezcla bien (la operación implica una mezcla aleatoria genuinamente mecánico-cuántica) y se extrae a ciegas el billete premiado. El procedimiento es el mismo en ambas loterías, pero en la lotería B el sorteo se realiza *antes* de la venta de los billetes. Son muchos los que opinan que la lotería B no es imparcial, ya que el billete ganador está determinado de antemano y los otros son simples papeles sin valor alguno cuya venta constituye una especie de fraude. Sin embargo, ambas loterías son igualmente imparciales; la posibilidad de ganar es la misma y el hecho de que el billete premiado se seleccione antes o después no tiene la menor importancia.

En realidad, si el sorteo se pospone hasta después de la venta de los billetes es sólo para que el público tenga un testimonio fehaciente de la honestidad del procedimiento. No cabe la sospecha de que algún agente (interiorizado en el funcionamiento de la lotería) manipule la distribución de los billetes, pues el conocimiento de cuál será el premiado no es accesible al agente hasta que no se vendan todos los billetes. (En efecto, este método se lo ha diseñado para demostrar que la lotería es a prueba de ogros.) Es interesante observar que no todas las loterías siguen el mismo procedimiento. El *Publishers Clearing House y el Reader's Digest* envían anualmente millones de sobres con la siguiente leyenda, escrita en grandes caracteres negros: "Usted ya puede haber ganado" un millón de dólares o cualquier otro premio. Estas costosas campañas se basan, seguramente, en investigaciones de mercado que demuestran que, para la opinión general, las loterías cuyos ganadores se seleccionan de antemano son imparciales, siempre y cuando se las maneje con honestidad. ¿Cuántas personas comprarán un billete de lotería sabiendo que el talón premiado yace, desde un principio, en la bóveda de un banco, dentro de un sobre sellado? Sospecho que la mayoría de las personas no se inmuta ante tal disposición, considerando que tiene una oportunidad real de ganar; es decir, sospecho que las personas comunes son menos supersticiosas que los filósofos (volviendo a Demócrito y a Lucrecio) quienes se han convencido a sí mismos de que, sin un suministro inagotable de *cántaros* que rompa la tela de la causalidad, no tendrán oportunidades o posibilidades reales.

Si nuestro mundo está determinado, ello significa que poseemos, en nuestro interior, generadores del número seudoaleatorio y no generadores del número aleatorio equipados con contadores Geiger. Es decir, que si nuestro mundo está determinado, todos nuestros billetes se han sorteado al mismo tiempo, desde hace eones, se han puesto en un sobre y se distribuyen a lo largo de nuestra vida, a medida que los necesitamos. Algunos aducirán que el método es injusto, pues ciertas personas recibirán más billetes que otras; además como el sorteo se efectúa antes del nacimiento, hay quienes están *destinados* a tener más suerte. Se debería recordar, en primer término, que en toda distribución hay quienes reciben cartas más altas que otros y, en segundo término, que ocurriría lo mismo aun cuando el sorteo no se efectuara antes del nacimiento sino periódicamente, a medida que lo requiriésemos. Incluso en un sorteo aleatorio, imparcial, genuinamente indeterminista, está *determinado* que habrá un ganador, como ocurrió en el concurso de cara o cruz. Lo que no puede afirmar el ganador es que "su destino" era ganar, pero (destino o no) todas las ventajas estaban de su

parte, y ¿puede haber algo más justo que eso? La justicia no consiste en el hecho de que todos ganen.

Hablé del concepto de *oportunidad* relacionándolo con el beneficio que comporta para los agentes o autocontroladores, porque éste es el territorio natural del concepto (en un oscuro sentido que no he podido elaborar en profundidad), si bien desempeña un papel mucho más vasto en la ciencia. Algunos filósofos lo han considerado así y han separado las cuestiones científicas del problema del libre albedrío y de los conceptos que le son afines:

> La oportunidad es un ejemplo notable de un concepto estrechamente vinculado con la *posibilidad*, pero que no puede tener sino una aplicación metafórica descabellada fuera de la esfera personal. (Ayers 1968, pág. 9.)

Pero esto no es así. Considérese, por ejemplo, el uso enteramente metafórico del concepto en Dawkins:

> La diferencia que deseo subrayar es la que existe entre la división celular en el caso de la célula germinal (reproducción) y la división celular en el caso de la célula somática o de "final definitivo" (crecimiento). En el primer caso, los genes, al duplicarse, tienen la *posibilidad* [la bastardilla es mía] de convertirse en los ancestros de una línea indefinidamente larga de descendientes, en donde los genes son, en realidad, verdaderos replicantes de la célula germinal... (Dawkins 1982, pág. 255).

Tampoco cabe descartar el uso del concepto cuando nos referimos a la posibilidad lógica o a la posibilidad mecánico-cuántica. Las oportunidades desempeñan un papel primordial no sólo en la concepción del mundo biológico sino también en la del mundo inanimado de la física y de la química, como veremos en el capítulo 6. Pero lo que nos interesa aquí son las oportunidades humanas; esto es, la posibilidad de elegir sobre la base de una información correcta. A mi entender, el problema de tener oportunidades, reales o no, es independiente del problema acerca de la verdad o la falsedad del determinismo.

La idea de que nuestras posibilidades serán sólo aparentes si el determinismo es verdadero no sólo está mal fundamentada metafísicamente sino que es políticamente retrógrada y moralmente ambigua, de una manera que no cabe ignorar. Si no hay oportunidades reales, tampoco habrá diferencias reales, importantes, entre las personas. Aquel que se angustia porque aún ignora si lo admitirán en Harvard, Stanford o Swarthmore estará en el mismo plano que el que *no se angustia* ante la posibilidad de concurrir o no al trabajo

en la fábrica. Cabe preguntarse cuál de los dos agentes tiene oportunidades *reales*. Pero ¿cuál de los dos agentes preferiría ser una persona cuerda? No promovemos la causa del libre albedrío pensando que la cuestión última sea imponderable y depende de las abstrusas implicaciones y de los descubrimientos de la física subatómica. Las oportunidades que interesan a las personas dependen de los elementos de su imagen manifiesta que no van más allá de lo cotidiano, y hay quien tiene más oportunidades que otros.

Por cierto, eso no tiene que ver con el lugar que ocupan en la escala socioeconómica. La falta de libertad, el "tener las manos atadas", no es exclusivo privilegio de los pobres o de los que no obtienen la carta de ciudadanía. Considérese el presidente de un banco, que no puede siquiera imaginar la idea de renunciar a su cargo y convertirse en jardinero (algo que le encantaría), aunque el trabajo lo esté literalmente matando a fuerza de úlceras e hipertensión. Si lo hiciera, su mujer encanecería de la noche a la mañana o bien enloquecería de furia; sus hijos tendrían que abandonar la universidad; sus colegas se sentirían desolados, sin atinar a explicarse tamaña decisión. Nosotros, que no estamos atados por esos lazos de afecto o compromiso, podemos burlarnos del dilema del banquero, pero ¿quién de nosotros sería capaz de una ruptura tan radical? Hay mucho de verdad en la canción cuya letra escribió Kris Kristofferson: "La libertad es sólo una palabra que se usa cuando no queda nada que perder" ("Yo y Bobby McGee"). Los que son sabios y afortunados, no importa el lugar que ocupen en el espectro socioeconómico, "mantienen sus opciones abiertas". La libertad del vagabundo es una libertad auténtica que el presidente del banco podría razonablemente envidiar; el vagabundo posee algo que el esclavo asalariado no posee, aunque ambos carezcan de poder político o social. Todo esto que digo no pasa de ser un lugar común, pero vale la pena tenerlo en cuenta toda vez que formulemos esas supuestas preguntas apasionantes acerca de nuestro lugar en el esquema de las cosas.

4. *Evitar, evitable, inevitable*

Soy un hombre viejo
y he visto muchos problemas
pero la mayoría nunca sucedió.
—Mark Twain

El Deliberador Determinista Mark I nos instruyó acerca de la naturaleza de las posibilidades. Se puede objetar, sin embargo, que

si bien tiene oportunidades genuinas que generalmente aprovecha con criterio, sus decisiones están determinadas y son, por lo tanto, inevitables. Si el determinismo es verdadero, todos nuestros actos serán inevitables, pues son los efectos de acontecimientos del pasado que se hallan fuera de nuestro control.

Esta manera de argumentar es tan común como incoherente y comporta un desplazamiento ilegítimo de "lo determinado" o "causalmente necesario" a "lo inevitable". Son muchos los autores que no reconocen la ilegitimidad del desplazamiento. Ayers (1968, pág. 5) se refiere a lo que es "necesario e inevitable". Wiggins (1973, págs 42-43) propone abreviar la modalidad "causal o histórica de los tiempos verbales" por medio de la expresión "inevitable en *t*", y a causa de una abreviatura incurre en todas las connotaciones que implica lo inevitable. Chisholm (1961, pág. 158) defiende (como uno de los miembros del dilema) el supuesto de que "los actos no son evitables si poseen suficientes condiciones causales". Veremos luego que la afirmación, evidente en apariencias, no pasa de ser una tautología inofensiva e irrelevante para nuestros intereses como deliberadores. Sin embargo, la supuesta peligrosidad reside en que:

> si el determinismo es verdadero, no podremos evitar ninguno de nuestros actos (dado que todos ellos tienen suficientes condiciones causales).

Pues cabe pensar que nos está permitida la siguiente inferencia: si el determinismo es verdadero, ningún agente podrá hacer nada respecto de nada. Pero eso es fatalismo y no se debe confundir fatalismo con determinismo.

No nos sorprende que a pesar de la profunda atención dispensada a la lógica modal de la necesidad o a la llamada inevitabilidad histórica, se le haya prestado tan poca atención al verdadero significado del término inevitable (Slote, 1982). Cuando se dice que algo es inevitable *tout court*, significa que no era *evitable* para alguien o, en el caso defectivo, que no es evitable para nadie. Pero si deseamos saber qué significa inevitable para todos, hemos de elucidar, en primer término, lo que significa decir que algo es (o era) evitable; y para esclarecerlo, se requiere saber en qué consiste evitar algo.

"Evitar" pertenece a la familia de los verbos que usamos al referirnos a los objetos y resultados de las deliberaciones: "prever", "anticipar", "impedir", "eludir", "realizar", "prevenir", "asegurar", "proteger", "preservar", "depender", "ayudar", "fomentar", "frustrar", etc. Estos son los principales verbos relativos a la agencia y los que usamos para describir los actos de aquellos que "cambian el curso

de la historia", "perturban el universo" o "establecer una diferencia". Quizá nuestra aspiración más secreta sea la de "establecer una diferencia".

Pero si nuestro deseo es el de modificar el curso de la historia, estamos condenados a sufrir un gran desengaño, pues nadie puede modificar su curso (por razones que nada tienen que ver con el determinismo). A comienzos del capítulo, imaginamos la totalidad del espacio y del tiempo desplegada ante nuestra mirada *sub specie aeternitatis*, para decirlo en jerga filosófica. Si se supone que la escena imaginada es el curso *real* de la historia a través de la eternidad, entonces —compruébelo usted mismo— la imagen carecerá de bifurcaciones. Como solamente puede aontecer una cosa real,[16] esté o no determinada, la parte de la imagen que lleva el rótulo de "futuro" estará compuesta por los acontecimientos que ocurran en la totalidad del tiempo. De esta manera, nuestra imagen de un universo indeterminista debería ser indiscernible de la imagen de un universo determinista en el tiempo. El futuro de hoy es la historia de mañana: la secuencia de los acontecimientos que suceden realmente son el eterno presente verbal.

¿En qué consiste cambiar el curso de la historia? Específicamente ¿qué parte de la historia? ¿Se trata de reemplazar un acontecimiento del futuro por otro? Presentada en esta forma, la idea resulta harto incoherente. Se suele decir que nadie puede modificar el pasado, pero raras veces se agrega que tampoco se puede modificar el futuro. Si el pasado es inmodificable, el futuro es inevitable en lo que a nosotros respecta. El futuro, independientemente del tiempo, es la secuencia de acontecimientos que van a suceder, estén o no determinados, y tiene tan poco sentido hablar de evitarlos como de evitar los acontecimientos que ya sucedieron. ¿Qué significa evitar algo? Hablamos de evitar *cosas* (las abejas evitan el árbol apartándose de él) pero, teniendo en cuenta nuestros objetivos, hablamos especialmente de evitar acontecimientos, máxime si los consideramos lamentables: calamidades, desastres, desgracias. Palabras como "soluble", "inflable", "frágil" y otras del género nos permiten señalar casos reales en los que algo que tiene (o tuvo) esa propiedad se disuelve se infla o se rompe. Pero no podemos señalar acontecimientos reales que se hayan evitado. Nada de lo que sucedió, suceda o sucederá puede ser un acontecimiento que fue, es o será evitado. Si lo que se desea son ejemplos, he aquí unas cuantas catástrofes que logramos evitar: el gran terremoto de Lon-

[16] A menos que adoptemos la "interpretación de los múltiples mundos" de la mecánica cuántica propuesta por Hugh Everett. Véase Hofstadter y Dennett 1981, págs. 44-46.

dres del año 1981; los disturbios y la masacre que se produjeron en el desfile inaugural de la campaña de Barry Goldwater en 1963; el gran terremoto de Londres de 1982. De acuerdo, se trata de acontecimientos imaginarios que obviamente nunca ocurrieron. Pero, entonces, ningún acontecimiento que suceda realmente, ningún acontecimiento, determinado o no, que pertenezca al conjunto de los acontecimientos futuros podrá ser evitado.

En el capítulo 3 vimos que no son las "fuerzas causales" que se encuentran fuera de nuestro control sino las fuerzas causales *inesperadas*, o bien esperadas pero imposibles de dominar, las que amenazan el control de una situación en particular. Esta última clase de fuerzas son las que nos proporcionan el modelo —escalofriante— de la *inexorabilidad*. Supóngase que los astrónomos descubren la existencia de un enorme cometa, y luego de calcular su trayectoria, afirman que caerá exactamente dentro de una semana en los Estados Unidos y *que no hay nada que se pueda hacer al respecto.* Ante semejante perspectiva, la población, además de ponerse frenética, imaginará toda clase de estrategias, preguntándose una y otra vez si no habrá *algo* que pueda hacerse. Y terminará de rodillas, suplicando un milagro. Supóngase que en la víspera de la catástrofe aparece otro cometa que desvía al del Juicio Final de su fatídica trayectoria, *impidiendo* la muerte de millones. ¿Se trata de un milagro? No. El segundo cometa no salió milagrosamente de la nada sino que siguió una ruta preestablecida durante millones de años. Los astrónomos no habían reparado en su existencia, pero revisando viejas fotos pertenecientes a los archivos de los observatorios se demostró que allí estaba, en camino para salvarnos. Luego, nunca existió ninguna emergencia. Si los astrónomos hubieran tenido mejores datos o utilizado mejor los que ya tenían, no hubiesen dado el alerta.

El segundo cometa nunca impidió la catástrofe por la simple razón de que nunca existió la posibilidad de una catástrofe. También se podría aducir, en ese sentido, que innumerables fuerzas astronómicas se aúnen cada día (o cada milésimo de segundo) para impedir la espantosa catástrofe que provocaría el choque de Marte contra la Tierra. La colisión no figura en los mapas, como tampoco figuraba la del temido cometa, a pesar de lo que se pensó en un principio. Lo que hizo que la catástrofe pareciera inminente y evitada por milagro fue, en realidad, el error cometido por los astrónomos al *prever* una cierta trayectoria en el futuro. Y, comparados con ese futuro erróneamente anticipado, los acontecimientos ulteriores se consideraron como una instancia de prevención.

Los verbos relacionados con "establecer una diferencia" conllevan todos una comparación tácita entre *lo que se supone* que

será la marcha del mundo y lo que ésta resulta en la realidad.[17] "Cambiar el curso de la historia" significa que los actos de un agente contribuyen de manera *visible* y fundamental a la marcha real del mundo —comparada con lo que se anticipó (o previamente se juzgó) que sería esa marcha—. El uso del concepto pide algo semejante a un principio de inercia de lo normal y se basa en la existencia de un fondo de suposiciones tácitas respecto del acontecer de las cosas "al ser iguales a otras cosas". Cuando el acto de un agente, comparado con este fondo de suposiciones o anticipaciones, pertenece a la categoría de las cosas que son distintas que otras, "establecerá toda la diferencia".

Nuestros hábitos de expectativa (entendidos a la manera de Hume) están tan profundamente arraigados, que se necesitan saltos sustanciales de la imaginación científica para discernir lo que requiere explicación y lo que no la requiere. Solamente cuando Newton pudo preguntarse qué impedía que la manzana permaneciera vagando en el espacio o suspendida a mitad de camino, advirtió la necesidad de la existencia de una fuerza gravitatoria. Los únicos acontecimientos que requieren explicación en la física newtoniana son las aceleraciones. El resto se explica, como de costumbre, por el movimiento rectilíneo. Una vez ubicadas las fuerzas básicas, podemos alcanzar el umbral de lo que requiere una explicación más amplia. Nos preguntamos qué impide que una piedra, aparentemente sin apoyo, caiga con aceleración gravitatoria; no nos preguntamos, obviamente, qué le impide estallar en llamas o transformarse en una taza de té.

El concepto de prevención (y su familia) lo mismo que el concepto de oportunidad, se explica con más exactitud si evaluamos el papel que desempeña en las deliberaciones de las personas y otros autocontroladores, pero ello no significa que no tenga ninguna aplicación fuera del dominio de los proyectos humanos. Consideremos el siguiente diálogo, tan habitual como fatalista:

> —¿Por qué te tomas el trabajo de ponerle cerrojo a la puerta?
> —Para evitar que las personas no autorizadas entren, por supuesto.
> —Pero si está determinado que no entrará en tu casa ninguna de esas personas, podrías dejar la puerta abierta de par en par; y si no está determinado que no entrará a tu casa una persona no autorizada, la lógica más elemental te diría que un cerrojo no va a impedírselo.

[17] Un buen ejemplo de alguien que *casi* lo logra (pero pasando por alto el giro epistémico) es von Wright (1974, págs. 42 y sigs.).

Somos lo bastante cautelosos para no tener en cuenta el argumento, tan cautelosos como fue la Madre Naturaleza cuando diseñó e instaló en los organismos toda clase de dispositivos de prevención y evitación. ¿Para qué tomarse el trabajo de dotarlos de anticuerpos, párpados, espeso pelaje o piel coriácea? Simplemente, no podríamos comprender ningún aspecto del mundo viviente, evolucionado, sin la ayuda de cantidades considerables de prevención y evitación. Si el determinismo demostró que *nadie* podía impedir nada, lo hizo demostrando que no hay una onza de prevención en ninguna parte: *nada* puede impedir nada. El sonido de las cataratas del Niágara no le impide al pájaro que se halla en las inmediaciones escuchar los pasos del gato que se aproxima, ni la gravedad terrestre puede impedir que los átomos que componen la atmósfera se escapen al espacio exterior.

Lo primero que cabe observar en esta familia de conceptos es que se vincula, en cierta manera, con lo normal y esperable —esto es, esperable para los que tienen expectativas—. ¿Cuál es la diferencia entre el gran terremoto de Londres de 1982 y la desaparición de la isla de Three Mile? Ninguno de los dos ocurrió, pero el último acontecimiento estuvo a punto de suceder y *fue evitado*. ¿Cuál es la diferencia entre los disturbios acaecidos en la campaña inaugural de Goldwater y la condena de Lee Harvey Oswald? Ninguno de los dos ocurrió, pero el último fue impedido por Jack Ruby cuando intentó (con éxito) perturbar el universo. Tanto el juicio y la condena de Oswald como la desaparición de la isla de Three Mile eran acontecimientos previstos, que formaban parte del normal discurrir de las cosas.

Hablando con Roger Schank, éste observó que nuestra manera de usar el artículo determinado es terriblemente reveladora. No nos irritamos cuando nos dice: "El hombre se sentó a la mesa del restaurante y el mozo le alcanzó el menú", pero lo hacemos si nos dicen: "el hombre se sentó a la mesa del restaurante y *el* jorobado le pidió fuego". ¿*Qué* jorobado? (no preguntamos: "¿*qué* mozo?") Este uso del artículo revela nuestro *compromiso* mental con diversas entidades previsibles, pero no en el sentido usado por los filósofos cuando se refieren al "compromiso ontológico", sino como usan el término los entrenadores de baloncesto cuando aconsejan comprometer primero al contrincante antes de decidir el curso del propio juego.

Todo plan o deliberación requiere que se tenga en cuenta los elementos del mundo que continuarán "normalmente" y aquellos elementos que pueden cambiar, cambiarán o se planea cambiar. (Este hecho constituye el meollo del problema de la estructura de la inteligencia artificial. Véase Dennett, en preparación-a.) De ese modo, los proyectos y previsiones entran en la categoría de las cosas

que van a suceder, *a menos que* se tomen ciertas medidas, de las que sucederán *porque* se tomaron ciertas medidas o de las que sucederán, *a pesar de las medidas que se tomen.* A estas últimas de las llama "inevitables", pues no se pueden modificar, se haga lo que se hiciere; de ahí la inutilidad de las deliberaciones.[18]

"Inevitable" no significa "causalmente necesario" o "determinado"; tampoco estos términos lo implican. El tratamiento que los filósofos dan al problema adolece de una curiosa inestabilidad de perspectiva que cabe interpretar como el oscuro reconocimiento de este vacío. Cuando los autores discuten (y deploran) las implicaciones del determinismo, se muestran con frecuencia ambivalentes. En ocasiones, parecería que incluyen al agente y a sus deliberaciones en la gran trama de la causalidad considerada por el determinismo; en otras, quisieran excluir del mundo determinista al desventurado agente con el propósito de librar una pequeña parte de su historia de la telaraña determinista (Greenspan 1978).

Esta manera ambivalente de pensar no sólo se encuentra en la filosofía. Cuando los científicos realizan experimentos sobre la naturaleza, se colocan fuera del mundo y con total independencia penetran en él para arancar la trama de la causalidad. Después aclaran, entre paréntesis: "También somos parte de esa trama, qué duda cabe". Pero no se necesita una polaridad tan radical en la perspectiva, pues ambas visiones pueden fundirse en una sola. Para que los experimentos sobre la naturaleza que los científicos llevan a cabo sean epistemológicamente potentes, no necesitan ser absoluta y metafísicamente independientes de la corriente causal que están examinando, sino sumamente desvinculados de los elementos de esa corriente causal. Lucas expresó con claridad esta exigencia de absoluto:

> Debemos ser agentes y no sólo espectadores. La razón es la siguiente: como agentes libres creemos que cabe introducir una perturbación arbitraria en el universo para destruir cualquier armonía preestablecida. Entre las transformaciones provocadas por nuestras arbitrarias intervenciones, sólo se preservarán las regularidades verdaderas y se destruirán las preestablecidas y las coincidentes. (1970, pág. 63.)

[18] Berlin (1954) descalifica con sutileza la perniciosa idea de la inevitabilidad histórica, pero no aclara bien la distinción entre *fuerzas históricas inexorables* (como nuestro cometa o un maremoto) y las *fuerzas que determinan los acontecimientos históricos operando a través de las deliberaciones de los agentes involucrados.* Berlin parece no darse cuenta de que no necesita romper con este último supuesto, que es consistente con su vigorosa descalificación de esa nefasta forma de determinismo histórico: la inevitabilidad histórica.

Pero el objetivo de forzar las "verdaderas" regularidades en el mundo para que se muestren a sí mismas también se logra (como vimos en la discusión sobre los usos del desorden en el capítulo 3) toda vez que nuestras actividades exploratorias "voluntarias" no están inadvertidamente en fase con alguna pauta ni son rastreadas por ningún demonio. En tanto dependan de procesos caóticos o seudoaleatorios, darán ventajas exploratorias a quienes las usan —no una ventaja absoluta que garantice en todo momento que no se está en fase con lo que se investiga, sino una ventaja suficientemente buena para la ciencia humana—. Según la absoluta certeza cartesiana, nuestras actividades exploratorias no sólo no deberían ser rastreadas: tampoco deberían ser rastreables en principio. De ahí la apelación al perfecto azar por parte de los absolutistas (véase también Nozick 1981, págs. 10 y sigs., y von Wright 1974, págs. 39 y sigs.).

Cuando a la confianza en tales estrategias de ruptura de vínculos se le suma una deliberada ceguera respecto de los procesos causales que la razón dicta, basta una ínfima exageración para crear la ilusión de la "absoluta agenciedad". Por lo tanto, la posible exclusión del agente de la corriente causal general da origen a reflexiones fatalistas y éstas, a su vez, originan la falsa elisión entre el determinismo y lo inevitable. Para percibir el sabor amargo de lo inevitable, se debe imaginar el momento, no importa si fugaz, cuando uno percibe que toda deliberación es inútil. En ese instante uno "ve", horrorizado, que está cayendo, cayendo y observando: la propia máquina deliberativa se ha desenganchado y sus ruedas giran inútiles mientras el drama incesante e inexorable se desarrolla allá afuera. En este ámbito imaginario, el propio proceso mental se separa temporaria y *tácitamente* de la naturaleza, permitiendo al que imagina un punto de vista relativamente estable a partir del cual puede experimentar el desnudo fluir de la causación—la clase de desasimiento con la que se observa el fatalismo local de un estornudo o de una eyaculación inminente—.

Pero toda la imaginería sobre el fatalismo local está fuera de lugar en esta discusión. Es sólo en aquellas ocasiones relativamente raras de fatalismo local, cuando la propia maquinaria deliberativa se separa verdaderamente de las trayectorias que nos interesan. Habitualmente, el pensamiento presente, el pensamiento-hasta-el-último-momento es crítico hasta el fin del resultado, con respecto a la finalidad del resultado. Según Hobart,[19] "el pasado no determina el acontecimiento sino a través del presente" (Hobart 1934). Las

[19] "R. E. Hobart" era el seudónimo de Dickinson S. Miller (1868-1963).

depresivas imágenes del fatalismo local no pertenecen a la categoría de *lo que debe ser* un pensador determinista en un mundo determinista. Pero en tanto nuestro disgusto por el determinismo esté coloreado por tales imágenes, no pensaremos seriamente en el tema.

En este capítulo, comenzamos por analizar el supuesto de que si el determinismo fuera verdadero, la imagen manifiesta de las oportunidades que se aprovechan y de las tragedias que se evitan sería ilusoria. Sin embargo, las únicas ilusiones que encontramos fueron inducidas por la negligencia de los filósofos, por la falta de atención a los detalles y por el abuso de los generadores de intuición. La aparente incoherencia de un deliberador determinista desaparece tan pronto se lo examina con más cuidado. Todo el mundo cree, aparentemente, que si el determinismo es verdadero, no habrá oportunidades reales y que las habrá si el determinismo es falso, pero esto es también una ilusión.

A pesar de lo que parezca a primera vista, algo puede ser "causalmente necesario" pero no inevitable. No hemos solucionado todos los problemas ni respondido a todas las preguntas acerca de la relación entre nuestra imagen manifiesta y nuestra imagen científica, pero sí hemos visto que los fundamentos tradicionales que justifican la alarma respecto del determinismo son el producto de un examen apresurado y no de una visión profunda.

6

"Podría haber obrado de otra manera"

1. La importancia de haber podido obrar de otra manera

En el seno de las discordias y desentendimientos que se suscitan entre los filósofos a causa del problema del libre albedrío, existen, todavía, algunas islas en calma donde reina, generalmente, la unanimidad. Van Inwagen señala:

> Casi todos los filósofos coinciden en que la condición necesaria para afirmar que un agente es responsable de un acto es creer que el agente pudo abstenerse de realizar ese acto. (Van Inwagen 1975, pág. 189.)

Si esto es así, aún no he tocado el tema central del libre albedrío, pues hasta el momento no he declarado mi posición sobre el principio "podría haber obrado de otra manera"; es decir, el principio que sostiene que sólo se actúa libre y responsablemente cuando se podría haber obrado de un modo distinto. Es tiempo, entonces, de volver a este dominio fundamental y estable en la geografía del problema del libre albedrío. En primer término, comenzaré por demostrar que dicho principio es sencillamente falso; luego retomaré algunos problemas que quedaron pendientes sobre el significado del verbo "poder" —la rana de Austin en el fondo del jarro de cerveza (véase el capítulo primero, pág. 19).

El principio de "podría haber obrado de otra manera" fue discutido por generaciones de filósofos. Los compatibilistas, para quienes el libre albedrío y el determinismo son, después de todo, compatibles, tienen una estrategia favorita: sostener que dicho principio no significa lo que parece a primera vista. El sentido de la

frase negado por el determinismo es irrelevante para el sentido que requiere la libertad. Pero es tan obvio que *tienen* que decir una cosa así, que son muchos los escépticos que rechazan estos análisis por considerarlos una suerte de alegatos destinados a engañar a los mismos compatibilistas. James (1921, pág. 149) denominó al tema "una charca de evasión" y Kant en *La crítica de la razón práctica*, lo calificó de "subterfugio lamentable".

Pero no voy a defender los análisis compatibilistas, muchos de los cuales me parecen absolutamente defendibles, sino que continuaré, más bien, en la ofensiva. En primer lugar, el significado de dicho principio no es lo que nos interesa cuando intentamos esclarecer si un acto se realizó libre y responsablemente. Según Van Inwagen, se ha convertido en una especie de tradición suponer que son acertadas las intuiciones en favor del principio; pero los filósofos que así lo suponen, lo hacen a pesar de los evidentes y bien conocidos argumentos que fundamentan la duda.

Uno de los pocos filósofos que recusan esta postura es Frankfurt, inventor, además de un interesante y productivo generador de intuiciones, cuya virtud consiste en suministrar ejemplos de todos los colores en contra de la posición tradicional: casos en donde el exceso de determinación asegura que si un agente, luego de haber elegido a sabiendas realizar un acto, se arrepiente de ello, entonces el infaltable ogro se encargará de que lo haga de todas maneras (véase Frankfurt 1969 y también Van Inwagen 1978 y 1983, y Fischer 1982). El generador de intuiciones que presentamos a continuación es un generador básico, desnudo, sin ninguna clase de aditamentos (que no nos interesan... porque no confiamos en ellos):

> Jones odia a Smith y decide, en plena posesión de sus facultades mentales, asesinarlo. Black, el neurocirujano nefando, también desea la muerte de Smith. Por lo tanto, implanta un dispositivo en el cerebro de Jones. Pero, *sólo si Jones cambia de parecer* (y se acobarda), Black apretando un botón pondrá nuevamente a Jones en la huella asesina. Black no interviene en el acontecimiento. Jones realizará el acto por sí solo.

Frankfurt afirma que en este caso la persona será responsable de su acto, puesto que lo eligió luego de las debidas deliberaciones y por voluntad propia, a pesar de la presencia oculta de Black, que determina (en exceso) que Jones no pueda actuar de otra manera.

Acepto los análisis de Frankfurt, es decir, creo que cabe defenderlos contra las objeciones de Van Inwagen, Fischer y otros autores; pienso, también, que estos experimentos mentales son útiles pues subrayan la importancia, respecto de la responsabilidad, de la cadena causal de deliberaciones y elecciones que funciona

en el agente, no importa lo que pueda estar sucediendo en otra parte. Pero la estrategia de Frankfurt me parece poco ambiciosa. Aunque utiliza sus ejemplos para demostrar que el famoso principio —que denomina "principio de las posibilidades alternas"— es irremediablemente falso, sus ejemplos son un tanto especiales y poco verosímiles e invitan a los defensores del mencionado principio a modificarlo ligeramente, de manera de protegerlo de los incómodos casos de Frankfurt. Por otra parte, las circunstancias exóticas no ayudan a disipar la ilusión de que, en el curso normal de las cosas, donde no existe tal exceso de determinación, el principio reinante sea, en verdad, el de no considerar responsable a la persona que no se pudo abstener de realizar un acto. Pero el hecho de que una persona podría o no haber actuado de otra manera no nos preocupa habitualmente; si en algunas ocasiones *parece* preocuparnos es porque, probablemente, deseamos sacar la conclusión opuesta acerca de la responsabilidad a partir de la conclusión que la tradición nos endilga.

"Aquí estoy, no puedo hacer otra cosa", dijo Lutero. Lutero afirmaba que no podía hacer otra cosa que aquello que, ante su conciencia, le era *imposible* desmentir. Por cierto, puede que estuviera equivocado o que exagerara deliberadamente la verdad. Pero incluso si así fuera —o, precisamente, si así fuera— su declaración testimonia que no debemos eximir de culpa o elogio a nadie por causa de un acto particular, simplemente porque pensamos que la persona no podía obrar de otra manera. Sea lo que fuere, una cosa es cierta: Lutero no estaba tratando de eludir ninguna responsabilidad.

Hay casos en los que la frase "no podría haber hecho otra cosa" significa una confesión de debilidad. Supongamos que debo volar en avión, pero permanezco clavado en el suelo, confesando que no puedo hacer otra cosa por culpa de mi miedo de volar, irracional e inhabilitante. Es decir, no puedo hacer otra cosa porque mi facultad de control racional está deteriorada. Pero, en otras ocasiones, como en el caso de Lutero, cuando digo que no puedo hacer otra cosa significa que no puedo hacerlo porque veo con claridad meridiana cuál es la situación y porque mi facultad de control racional *no está* deteriorada. La razón me dicta lo que debo hacer y tendría que estar loco para actuar de otra manera; luego, como no estoy loco, no puedo actuar de otra manera. (Obsérvese que "incumbía" a Lutero desdecirse o no, y no nos sentimos tentados a anular ese juicio, sobre todo cuando nos enteramos de que afirmó "que no podía hacer otra cosa". Obsérvese, también, que a menudo usamos expresiones como ésta: "Si fuera de mi incumbencia, yo sabría con certeza lo que tengo que hacer".)

Pienso que nadie en este mundo podría inducirme a torturar a una persona inocente por 1.000 dólares (y espero que sea cierto, y pienso que probablemente lo es). Pero la objeción no tarda en llegar: "¿Qué pasaría si ciertos piratas malignos del espacio tuvieran en sus manos el destino del planeta y prometieran no destruirlo sólo si usted acepta torturar a un inocente? ¿Seguiría pensando lo mismo?". Probablemente no, pero ¿qué importancia tiene? Se supone que uno está interesado en saber *si en circunstancias específicas* pudo haber obrado de otra manera; en consecuencia, la objeción es absolutamente irrelevante. Afirmé que sería imposible inducirme a torturar a una persona inocente *por 1.000 dólares*. Los que adhieren al famoso principio insisten en que uno debería discernir si podría haber obrado de otra manera *exactamente* en las mismas circunstancias. Yo afirmo algo más categórico: afirmo que no se podría hacer otra cosa incluso en un caso más o menos similar. No accedería jamás a torturar a una persona por 1.000 dólares. No importa qué tono de voz emplee quien intente sobornarme, que me sienta o no hambriento o que la víctima propuesta esté iluminada o se halle parcialmente en la sombra: soy (espero) inmune a tales ofertas.

Sin embargo, alguien podría aducir, basándose en sus intuiciones, que si estoy en lo cierto y si por casualidad se me presenta la ocasión de rechazar el ofrecimiento, mi rechazo podría considerarse como un acto no responsable. Puede que muchos piensen —cuando afirmé que me sería imposible actuar de otra manera en esa circunstancia— que soy una especie de *zombie* "programado" para rechazar un soborno de mil dólares cuantas veces se me presente la ocasión. Opinan que un auténtico agente libre debe ser un poco más voluble. Si soy capaz de escuchar la voz de la razón y de ser flexible de la manera apropiada, ¿por qué ser tan dogmático? Aun en los casos más extraños, debería considerar "las dos caras del asunto"; es decir, tomarme a mí mismo y sopesar los pro y los contra de la lucrativa tortura. Pero la única manera en que yo podría estar estructurado para considerar siempre "las dos caras" —aunque una de ellas fuese absurda— es estar estructurado de modo que *en cada caso particular*, "yo podría haber actuado de otra forma".

Este razonamiento es falaz. Considerar ambos lados de una cuestión no significa que uno no sea finalmente persuadido por uno de ellos. La flexibilidad que, en nuestra opinión, debe poseer un agente responsable es la de reconocer que en este caso (cuyas probabilidades son extremadamente remotas) se pudo salvar el mundo (o lo que fuere) gracias a esos 1.000 dólares imposibles de obtener de otra manera. Pero la capacidad general de dar respuestas flexibles en tales casos, no tiene por requisito "que uno podría

haber obrado de otra manera" *en el caso particular*, sino que uno hubiera obrado de otro modo solamente en las variaciones de las circunstancias —es decir, en las variaciones que importan—. Nos será útil considerar dos casos, que si bien parecen disímiles en un primer momento, pertenecen, sin embargo, al mismo *continuum*.

1. Supongamos que no ignoro que si veo la luna llena, probablemente me venga el *amok* y mate a la primera persona que encuentre. Por lo tanto, algunas noches al mes tomo precauciones para quedar encerrado en una habitación sin ventanas y me *inhabilito* entonces para no cometer las cosas terribles que, de no ser así, cometería. Además, gracias a mis esfuerzos responsables, me he vuelto *incapaz* de incurrir en tales acciones funestas. Se trata, sin duda, de un caso fantástico; consideremos el siguiente, bastante más realista.

2. Supongamos que no ignoro que si veo a una mujer voluptuosa caminando sola por un lugar desierto, probablemente me domine la lujuria y la viole. De modo que me instruyo acerca de los horrores de la violación, considerada desde el punto de vista femenino, y mi imaginación exacerba a tal punto la brutalidad del crimen. (Sin embargo, cada vez que me encuentro con una mujer que cumple esas condiciones, experimento un tumulto interior considerable y sacudo los barrotes de la jaula que yo mismo me he construido, lo que me convence de que, de no ser así, podría haber hecho esa cosa repudiable.) Gracias a mis esfuerzos responsables me he vuelto inmune a esta forma (bastante común) de poseer a las mujeres. Hice lo que debía para que ciertas acciones fuesen *impensables* en lo que a mí concierne. Y, como Lutero, *ahora* no puedo hacer otra cosa.

Supongamos —ahondando en el realismo— que nuestros padres y maestros no ignoran que sin una educación moral nos convertiremos en adultos egoístas, indignos de confianza y probablemente peligrosos. De manera que se las ingenian para educarnos y, gracias a sus esfuerzos responsables, podemos apartar de nuestra mente todo pensamiento de latrocinio, engaño o violencia. Estas alternativas nos resultan impensables en circunstancias morales; además nos enseñaron a cuidarnos a nosotros mismos y a contribuir a nuestro desarrollo moral —¿acaso ser una persona responsable no consiste, en buena medida, en inhabilitarse para cometer acciones que *nos volverían culpables?*—. Los filósofos señalan a menudo (con bastante incomodidad) que los problemas morales difíciles, las decisiones "que podrían haber tomado otro rumbo", no son los únicos ni los más frecuentes tipos de decisión que hacen que se considere responsable a un agente. Son raras las ocasiones en las cuales los filósofos sospechan o se preguntan si dicho principio no será, sencillamente, falso.

Acepto que se formule la pregunta: "¿podría haber obrado de otra manera?" luego de cometer un acto lamentable. ¡Pero nunca se ha demostrado interés alguno en responder a lo que presumiblemente se acaba de preguntar! Los defensores del principio suponen que "podría haber obrado de otra manera" tiene un sentido según el cual nadie podría hacer otra cosa que la que hizo, en caso de ser verdadero el determinismo. Supongamos que estén en lo cierto al afirmar la existencia de tal sentido. Ahora bien, cuando comenzamos una indagación acerca de la responsabilidad que le cupo a un agente por un acto que cometió, ¿nos proponemos usar la frase "podría haber obrado de otro modo" en ese sentido? No. Cuando nos dedicamos a ese tipo de indagaciones, ignoramos abiertamente la clase de investigación que debería realizarse si estuviéramos realmente interesados en responder la pregunta; es decir, a la pregunta metafísica acerca de si el agente está o no completamente determinado, al ejecutar una acción, por el estado del universo en ese momento.

Si, como afirma esta importante corriente filosófica, nuestra responsabilidad depende de la pregunta "¿podríamos haber obrado de otra manera *exactamente* en aquellas circunstancias?", nos enfrentamos con un problema de ignorancia de lo más peculiar, pues es hasta improbable, dado lo que al parecer establece la física, que se pueda determinar si un agente ha sido o no responsable. De acuerdo con la ortodoxia actual, reina el indeterminismo en el nivel subatómico de la mecánica cuántica; de modo que a falta de un argumento general y aceptado para el determinismo universal, es posible, hasta donde sabemos, que nuestras acciones y decisiones sean los efectos aumentados, macroscópicos, de indeterminaciones en el nivel cuántico que ocurren en nuestro cerebro. Pero también es posible, hasta donde sabemos, e incluso si el indeterminismo reina en nuestro cerebro en el nivel mecánico—cuántico, que nuestras decisiones y actos macroscópicos estén determinados. Los efectos cuánticos podrían anularse a sí mismos en lugar de amplificarse (como si los anularan los contadores Geiger orgánicos de las neuronas). Y es extremadamente improbable, dada la complejidad del cerebro aun en el nivel molecular (una complejidad que va más allá del predicado "astronómica") que podamos probar de manera contundente que un acto particular es el efecto en gran escala de una indeterminación en el nivel subatómico crítico. En consecuencia, si la responsabilidad de alguien por un acto depende de que la decisión esté o no determinada, en el momento de decidir, por un estado anterior del mundo, entonces, dejando de lado el regreso victorioso del determinismo universal en la microfísica (que aboliría toda responsabilidad en ese aspecto) es muy posible que

nunca tengamos ninguna razón para creer que un acto particular fue o no el responsable. La diferencia crítica sería completamente inescrutable desde toda posición de ventaja macroscópica, y prácticamente inescrutable desde toda posición de ventaja macroscópica, y prácticamente inescrutable desde la más sofisticada posición microfísica que pueda imaginarse.

Algunos filósofos podrían sentirse bien con esta conclusión, pero sospecho que *solamente* un filósofo puede sentirse así. Es muy difícil pensar seriamente en la idea de que algo que nos importa tanto trascienda hasta ese extremo nuestra capacidad cognitiva. (Pero cabe considerar el punto de una manera diferente: quienes dicen saber que podrían haber obrado de otro modo exactamente en aquellas circunstancias deben admitir que proclaman este hecho, presumiblemente empírico, sin el menor atisbo de evidencia y sin la más remota esperanza de obtener tal evidencia.) [1]

Dada la absoluta imposibilidad de llevar a cabo una investigación significativa sobre la pregunta ¿qué piensan entonces las personas cuando la formulan en los casos particulares? Puede que piensen que están formulando otra pregunta. Y no se equivocan: están formulando una pregunta mucho mejor. (Si unos pocos continúan todavía planteando la pregunta metafísicamente imposible de responder, es porque son inducidos a engaño por la filosofía.) La pregunta que a las personas les interesa formular es mejor por dos razones: se la puede responder empíricamente y la respuesta es importante. La pregunta metafísica tradicional no sólo es imposible de responder: la respuesta, si la hubiera, sería inútil.

¿En qué nos beneficia saber, de un agente particular, que en alguna ocasión (o en toda ocasión) podría haber obrado de otra manera o que no podría haberlo hecho? Comenzaremos por la instancia negativa. Supongamos que usted sabe (porque Dios se lo dijo) que cuando Jones apretó el gatillo y asesinó a su mujer en el tiempo t, no habría podido actuar de otro modo. Es decir, dado el microestado de Jones en t y el microestado completo del entorno de Jones en t (incluyendo los efectos gravitatorios de las estrellas más remotas, etc.), el único camino posible para Jones era el camino que tomó. Si volviera a poner a Jones exactamente en el mismo estado y exactamente en aquella circunstancia, volvería a apretar el gatillo; y si la operación se repitiera millones de veces, Jones gatillaría millones

[1] Raab (1955) afirma que la pregunta metafísica sobre la "ausencia de causalidad" es "indemostrable" y señala la peculiaridad de pensar seriamente en una pregunta que no tiene respuesta. La declaración de Raab se basa en la afirmación, sin duda verdadera, de que ningún agente puede saber (ni tener acceso privilegiado a ese conocimiento) si una acción tenía o no causas. A mi juicio, ninguna investigación podría esclarecer el punto.

de veces. Ahora bien, usted se ha enterado del hecho, pero ¿se enteró de algo acerca de Jones? ¿Se enteró, por ejemplo, de su carácter o de su probable comportamiento en situaciones similares? No. Aunque las personas sean objetos físicos que, como los átomos, las bolas de billar o los puentes, obedecen a las leyes de la física, no sólo son más complicadas que las cosas que conocemos en el universo, sino que además están diseñadas para ser muy sensibles al cambiante espectáculo que jamás se encuentra dos veces en el mismo microestado. No se necesita descender al plano subatómico para establecer esta verdad sencilla. Las personas aprenden, recuerdan, se aburren, prestan atención a diversas cosas y cambian sus intereses de una manera tan incesante que es infinitamente improbable que una persona se encuentre en el mismo estado *psicológico* o *cognitivo* en dos ocasiones. Y esto aun sería cierto, si lográramos que el entorno que la circunda fuera absolutamente el mismo en dos ocasiones diferentes, pues la segunda vez el agente pensaría, sin duda, algo que no pensó la primera; por ejemplo: "¡Oh, Dios, todo esto me es tan familiar! Ahora, veamos ¿qué hice la última vez?" (véase el capítulo 2, pág 33).

Nos interesa determinar cómo reaccionará un puente ante circunstancias muy específicas, puesto que son esas circunstancias las que podría encontrar *en el estado presente* en una ocasión futura. Pero acaso no sea relevante, en lo que al conocimiento se refiere, determinar la microcausalidad del comportamiento de un ser humano en una circunstancia determinada, pues no volverá a enfrentarse con la misma circunstancia, y si lo hiciera, se encontraría en un estado reactivo sustancialmente distinto, en ese momento. Saber (porque Dios lo dijo) que un agente *no* estaba determinado para actuar en esos términos, sería igualmente ocioso en lo que respecta a la valoración de su carácter o a sus proyectos futuros. Como vimos en el capítulo cinco, el agente indeterminista no será más flexible, versátil, sensible a los matices ni más perfectible que el agente determinista.

En consecuencia, el interés por saber si se podría haber obrado de otra manera, exactamente en las mismas circunstancias y en el mismo estado interno, supone la existencia de una curiosidad metafísica particularmente pura, tan pura que carece de todo motivo ulterior, ya que la respuesta no modificará en nada el transcurrir del mundo. [2]

[2] Nozick (1981, pág. 313) afirma que todos deseamos un valor "originativo", pero las únicas condiciones en las cuales podríamos obtenerlo son aquellas que aparentemente requieren una lectura metafísica de la frase "podría haber obrado de otra manera". "Queremos que sea verdad que en esa misma situación podríamos haber obrado

¿Por qué existe una diferencia tan abismal, en cierto modo visible desde el punto de vista humano, entre un mundo donde se podría hacer otra cosa y un mundo donde ello no sería posible? ¿Por qué es tan aterrador el determinismo? Acaso nos engañe la imagen omnisciente, *sub specie aeternitatis*, desde la cual atisbamos nuestra vida entera, desplegada en el espacio y en el tiempo desde el nacimiento hasta la muerte, como si se tratara de un único, inmutable, cuatridimensional "gusano de espacio—tiempo" clavado en la trama de la causalidad e incapaz de todo movimiento. (La causalidad, como dijo metafóricamente Hume, es "el cemento del universo" [Mackie 1974]; quizá por esa razón imaginamos nuestra vidas *vaciadas en cemento*, atrapadas como un fósil en la piedra inalterable del espacio—tiempo.)

Deseamos que nos demuestren que podemos *movernos* en ese ámbito. Pero hay aquí una confusión; experimentar ese deseo es haber olvidado que el tiempo es una dimensión que se vuelve espacial en nuestra imagen. Atisbar de izquierda a derecha es atisbar del pasado al futuro, y un corte vertical de nuestra imagen captura un instante único del tiempo. En ése ámbito, la libertad de acción —la libertad para arrastrarse y serpear entre los puntos fijos del nacimiento y de la muerte— no reside en poseer la facultad de elegir de manera indeterminada sino en la de elegir dos o más cursos de acción *en un solo tiempo*.

¿Deseamos, entonces, guardarnos el pastel y también comerlo? ¿Casarnos y permanecer solteros? ¿Gatillar y a la vez dejar caer el arma? Si nos referimos a esta clase de libre albedrío, entonces, valga o no la pena, nos resultará inalcanzable, *a menos que adoptáramos* la interpretación de los múltiples mundos de la mecánica cuántica de Everett; en ese caso, es dable deducir que hemos vivido un número incalculable de vidas, ¡aunque nuestros *alter ego* no se encuentren jamás para comparar las notas!

Pero absteniéndonos de toda fantasía y volviendo a la cordura, aún cabe una modesta esperanza: hay momentos en los que sacrificaríamos todo por regresar al pasado y reparar aquello que hicimos, pero el pasado es una instancia definitivamente cerrada para nosotros. Nos queda, entonces, el futuro y nos gustaría establecer "un futuro abierto". Un futuro abierto implica que mis deliberaciones serán efectivas y que si decido hacer A, haré A, y si decido B, pues

---

(significativamente) de otra manera, de modo que nuestras acciones tengan un valor originativo". Pero ¿es posible que nos importe tanto (si Nozick está en lo cierto) una cosa que nunca sabremos si es verdadera o falsa? Si éste es el requisito del "valor originativo" ¿por qué habríamos de preocuparnos por tener ese valor "originativo"?

entonces haré B; es decir, un futuro (pues sólo *un* futuro es posible) donde lo único que pueda suceder sea aquello que he decidido llevar a cabo, finalmente.

2. *Las cosas que nos interesan*

Puesto que no nos interesa saber si una persona podría haber obrado de otra manera (una vez que observamos macroscópicamente la causación involucrada), cabe plantearse cuál es la otra pregunta que nos interesa cuando preguntamos "¿podría haberlo hecho de otra manera?".

Usaré mi táctica habitual de responder, en primer término, a una pregunta más simple sobre entidades más simples —en este caso, el robot determinista Mark I—. Por hipótesis, Mark vivió su vida como una máquina determinista en un planeta determinista, de modo que todo lo que hace, no podría hacerlo de otra manera, si usamos los términos en el sentido estricto y metafísico que los filósofos les atribuyen. Supongamos entonces que un bello día marciano Mark I comete un error lamentable al proyectar y ejecutar un esquema, que destruye algo valioso —digamos otro robot—. Por el momento, supondré que Mark I no lamenta el hecho;[3] son los diseñadores que permanecen en la Tierra quienes lo lamentan, preguntándose, naturalmente, si Mark I podría haber obrado de otra manera.

Como se trata de un sistema determinista, las preguntas que se formulen no serán metafísicas sino que estarán referidas al diseño del robot. Luego del lamentable episodio, sería conveniente mejorarlo para evitar que este *tipo* de acontecimientos se repita en el futuro.[4] Los diseñadores desean conocer, ciertamente, la clase de información utilizada por el robot y la forma en que razonó o planeó el esquema; es decir, si razonó de manera correcta y, en ese caso, si razonó "lo suficiente". No ignoran que, en un sentido del término, el robot no lo hizo; de no ser así, habría hecho lo correcto. Además, si el robot hubiera usado óptimamente los procedimientos heurísticos óptimamente diseñados, no habría razones para mejorar el diseño

[3] Sólo porque en este caso he considerado un robot relativamente simple. Pero un robot que se hiciera a sí mismo como el yo se hace a sí mismo, según vimos en el capítulo 4, sería capaz de lamentar el hecho.

[4] "No estamos interesados en el funcionamiento de una máquina que manejela comunicación para un solo *input*. Para que funcione correctamente, debe desempeñarse satisfactoriamente para toda clase de *inputs* y esto significa un desempeño satisfactorio para la clase de *inputs* que estadísticamente se espera recibir" (Wiener 1948, pág. 55).

—pero las posibilidades heurísticas que manejó en esa ocasión no dieron resultado alguno—. Si el robot se encontrase en una situación similar en el futuro y gracias al hecho de que el generador del número seudoaleatorio se halla en un estado diferente, podría actuar de otra manera; en realidad haría lo correcto, como de costumbre. No caeré en la tentación de agregar que *podría* haber hecho lo correcto en aquella oportunidad. Ya que estaba bastante bien diseñado para obrar apropiadamente (pero el carácter de Mark I no está en discusión). El fracaso se debió a que algo que no estaba *diseñado* (y era imprevisible) intervino en el proceso, modificándolo en sentido negativo.

Ningún programa heurístico garantiza que el resultado será "exacto" o que suministrará el resultado que se le requiera. Algunos programas son mejores que otros y cuando fracasan, es posible atribuirlo a alguna imperfección propia del diseño. Pero incluso los mejores programas no son a prueba de tontos y a veces fallan sin razón aparente. Como diría Cole Porter: "Era sólo una de esas cosas".

Estas fallas no son las únicas que "cuentan", para los diseñadores, como casos en los que el sistema "podría haber obrado de otra manera". Si descubren que el desperfecto se debió a una "caprichosa" mota de polvo que al introducirse en determinado lugar desbarató el sistema, decidirán que es inútil rediseñar el robot a causa de un incidente cuya recurrencia es muy improbable.[5] En el caso microparticular, el robot no podría haber obrado de otra manera e incluso si se hallase, por la más remota de las casualidades, exactamente en las mismas circunstancias, fallaría nuevamente. No existe, pues, ningún interés racional en mejorar un diseño irreprochable, aunque fracasara en esa ocasión. Hay una gran diferencia entre estar óptimamente diseñado y ser infalible (véase el capítulo 7).

Consideremos otro caso. El robot posee un arma de rayos que dispara con un 99,9 por ciento de precisión; es decir, que, a veces, si las distancias son muy grandes, no da en el blanco. Cuando al robot le falla la puntería, los ingenieros se preocupan por averiguar las causas del yerro. ¿Se debió a un error *sistemático* en los controles, a una insuficiencia de desperfecto que puede ser recurrente o "era sólo una de esas cosas", esos "actos de Dios" en los que, a

[5] Literalmente hablando, la recurrencia de un acontecimiento *de este tipo general*; no hay necesidad de tomar recaudos para evitar la recurrencia de un acontecimiento particular (algo lógicamente imposible) o para evitar la recurrencia de un acontecimiento exactamente del mismo tipo (algo nomológicamente imposible).

pesar de la ejecución irreprochable de la rutina y del diseño óptimo, no se dio en el blanco por un estrecho margen? Siempre existirán casos de esta índole, pero el objetivo es reducirlos a un mínimo-compatible, desde luego, con la relación coste-eficacia—. Más allá de ciertos límites, no vale la pena preocuparse por los errores. Quine (1960, págs. 182 y 259) señala que los ingenieros tienen un concepto que supone algo más que un pasajero interés filosófico: el concepto de "no importa"; es decir, los casos que cabe ignorar si se desea proceder racionalmente. Cuando se aseguran de que un determinado error pertenece a la categoría de "no importa", se encogen de hombros y dicen: "Y bien, podría haber sido un éxito".

Cuando un robot funciona mal, lo que les interesa determinar a los ingenieros es si se trata de un mal funcionamiento significativo; esto es, si revelan un patrón de disfunción sistemática que puede ser recurrente o una relación inadecuada y poco favorable entre los tipos de circunstancias y los tipos de reacciones. ¿Es probable que vuelva a suceder esa *clase* de incidentes o se debió a una conjunción de factores independientes que es muy raro se repita? Para obtener tales evidencias, se ignoran los microestadios, que nunca volverán a ser los mismos en ningún caso y se analiza al robot dentro de un ordenamiento finito de estados *macroscópicamente* definidos, organizado de tal modo que existan vínculos entre los diversos grados de libertad del sistema. La pregunta que, en consecuencia, se formulan, es la siguiente: ¿son éstos los vínculos apropiados para la tarea?[6]

Este recurso, utilizado para ignorar el microdeterminismo (dondequiera exista "en principio") y para borrar suficientemente las sutiles distinciones dentro de los estados y regiones que se relacionan de manera probabilística y que cabe *tratar como* homogéneos, es un recurso claro, seguro y no problemático en la ciencia, especialmente en la ingeniería y la biología (véanse Wiener 1948 y Winsatt 1980). Ello no significa que ésta sea la única manera

[6] Shaler Stidham me señaló que en teoría de las colas [*queuing theory*], una rama de la investigación operativa, hay un método llamado técnica común de números aleatorios utilizada cuando se ejecutan simulaciones de sistemas de colas para ver cómo responden a la variación "aleatoria" en sus condiciones operativas. En la técnica se prueban diferentes disposiciones de los parámetros de diseño *en la secuencia misma de números seudoaleatorios*. Se trata, en efecto, de una investigación experimental en "mundos posibles", donde se prueban sistemas ligeramente distintos en *exactamente* los mismos mundos y se comparan sus funcionamientos. Para ciertos fines, se puede demostrar que esta prueba de fuerzas y debilidades relativas es más sensible que la simulación más realista, en la cual los distintos modelos son probados en conjuntos "microscópicamente" diferentes pero "prácticamente" indistinguibles o seudoaleatorios. En los trabajos teóricos de la teoría de las colas, la suposición de tales mundos aleatorios acoplados es conocida como la suposición de acoplamiento estocástico. Para una explicación introductoria véase Fischman (1973).

de pensar acerca de las personas cuando nos preguntan si actuaron con responsabilidad, aunque habría mucho que decir al respecto.

¿Por qué se formula la pregunta "podría haber obrado de otra manera"? Se la formula cuando se desea interpretar un acto en particular, comprendiendo el cómo, el porqué y el significado que se le debe atribuir. En suma, porque se quiere averiguar cuáles son las conclusiones que se pueden extraer de dicho acto con respecto al futuro. ¿El acto en cuestión, por ejemplo, dice algo acerca del carácter del agente? ¿No comporta una crítica que, presentada en los debidos términos, podría persuadirlo de mejorar sus actitudes en muchos aspectos? ¿Se puede inferir del incidente que la persona se comportará o no de una manera similar en circunstancias similares? ¿Si se mantuviera invariable el carácter del agente pero se cambiaran las circunstancias, atenuándolas o incluso agravándolas, cometería, la mayor parte de las veces, el mismo acto lamentable? ¿Lo que hizo, lo hizo "de carambola" o es la manifestación de una "sólida" tendencia que persiste o es constante aun en las condiciones más diversas? [7]

Cuando el agente en cuestión es uno mismo, la razón se vuelve claramente visible. Supongamos que he cometido un acto deleznable. *¿A quién* le importa si podría haber hecho otra cosa exactamente en las mismas circunstancias y en el mismo estado mental? No pude hacerlo y es demasiado tarde para reparar el daño.[8] Pero cuando interpreto lo que hice ¿aprendo algo acerca de mí mismo? ¿Debería practicar la maniobra que no atiné a realizar en aquella ocasión, con la esperanza de afianzarla y volverla menos vulnerable a las perturbaciones, o sería, gastar tiempo y esfuerzo? ¿Me beneficiaría corregir mis hábitos mentales a fin de no cometer tales actos en el futuro, si se presentara la ocasión?

[7] Nos interesan las carambolas y las tendencias en ambos sentidos (dignas de alabanza y dignas de repudio). Si tuviéramos la certeza de que Lutero sólo estaba bromeando y que su posición, aparentemente fiel a sus principios, era una especie de malentendido p coincidencia de ópera bufa, su fuerza moral se vería considerablemente menoscabada ante nuestros ojos. Al fin y al cabo, no era tan fiel como parecía y, diríamos, "podría muy bien haber obrado de otra manera".

[8] A veces me pregunto si una parte del atractivo subliminal de la "libertad de acción radical" o libertad "contracausal" no es el siguiente: "No puedo cambiar el pasado, pero me siento mejor conmigo mismo cuando pienso que *casi* lo habría cambiado"; *o bien* "me sentiría mejor conmigo mismo si me enterara de que, por una especie de descuido cósmico, *no* puedo ser responsable". Se ha afirmado, con frecuencia, que la responsabilidad y el determinismo son incompatibles (por ejemplo, Hobart 1934 y Ayers 1954). El argumento que habitualmente se esgrime es falaz, como demostré en Dennett 1978a, capítulo 15. Sin embargo, podría ser que esta presunta incompatibilidad entre la responsabilidad y la libertad "contracausal" sea la que atrae secretamente a muchos hacia la libertad radical o contracausal.

Dado que siempre estaré a merced del flujo de consideraciones que meramente me suceden mientras el tiempo prosigue su rápida carrera y que no puedo controlar ese proceso de deliberación, debo tomar medidas para evitar que ciertos pensamientos "acudan a mi mente" en situaciones críticas. Podría, por ejemplo, cultivar el hábito de contar hasta diez antes de emitir una opinión sobre Ronald Reagan, pues he aprendido que el tiempo de deliberación que así se gana, nos recompensa con creces, cortando de raíz el lamentable exabrupto que implica todo comentario poco moderado. También debería contar los vasos de vino que bebo cuando, sumergido en una conversación, no reparo en quien, solícitamente, inclina la botella hacia mi vaso. Esta vez me comporté como un tonto, de acuerdo. Si la situación hubiera sido diferente, seguramente habría obrado de otra manera. Si la situación hubiera sido virtualmente la misma, quizá podría haber obrado de otra manera o quizá no. Lo importante es tomar recaudos para actuar de otra forma cuando se presenten circunstancias similares en el futuro.

Por cierto, ésta es la actitud más sana a tomar respecto de los episodios lamentables que uno ha protagonizado en el pasado reciente; es la versión aplicada a uno mismo de la actitud del ingeniero hacia las imperfecciones en el diseño del robot. Naturalmente, si prefiero buscar excusas en lugar de mejorarme, puedo aducir que no *tengo* que asumir ninguna responsabilidad por mi acción, ya que siempre cabe imaginar un punto de vista más sutil, desde el cual la circunstancia se destaque más que yo y, por decirlo así, me exceda. (Si uno se vuelve muy pequeño, podrá externalizar prácticamente cualquier cosa.)

En el capítulo 7 investigaré con más detalle el fundamento de esta generosidad en la autodescripción de responsabilidades. Por el momento, deseo señalar un caso muy común: hay momentos en los que uno está a punto de preguntarse si podría haber obrado de otra manera, pero se abstiene (sabiamente) de formular la pregunta. Se suele decir, luego de cometer una acción deplorable: "¡Lo siento terriblemente! ¡Nunca imaginé las consecuencias! ¡No se me ocurrió que podía hacer tanto daño!". Esto se parece sospechosamente al comienzo de una excusa —"¿Acaso puedo evitar lo que me ocurre y lo que no me ocurre?"—, pero las personas sanas que se controlan a sí mismas se abstienen de transitar estos caminos. Se hacen responsables aun de aquello que probablemente no sea sino un "accidente". De ese modo, disminuye la probabilidad de convertirse, ellos mismos, en víctimas "accidentales" en el futuro.

3. El verbo "poder"

> *La posibilidad de los*
> *teóricos del* quantum, *no es la*
> *libertad ética de los agustinianos, y* Tyché *es*
> *una deidad tan implacable como*
> Ananké.
> — Norbert Wiener (1948, p. 49)

Estas reflexiones edificantes invitan a una embestida final escéptica: "Usted ha pintado un cuadro idílico de quienes se controlan a sí mismos y hacen todo lo posible por mejorar su conducta, mas ¿qué sentido tiene todo este *empeño*? Si el determinismo es verdadero, entonces lo que realmente sucede es lo único que *puede* suceder". En un mundo determinista, la exhortación a hacer todo lo posible carece de sentido. Sin embargo, vemos a diario a personas que hacen mucho menos de lo que podrían hacer. ¿Cómo debemos interpretarlo? Si el determinismo es verdadero y si ello significa que lo único que una persona puede hacer es lo que en realidad hace, entonces, sin intentarlo siquiera, cada persona hará siempre lo mejor —y también lo peor—. La exhortación es incomprensible, a menos que haya algún espacio entre lo real y lo posible y cierta libertad de acción para maniobrar en consecuencia. Y no sólo esto: el juicio y la valoración retrospectiva se vuelven aparentemente inútiles. No sólo no será verdad que cada uno hace siempre lo mejor que puede, sino que cada cosa será tan buena como puede —o tan mala—. El doctor Pangloss, famoso por su optimismo, estará en lo cierto: éste es el mejor de todos los mundos posibles. Pero su némesis, el pesimista doctor Pangloss, asentirá, suspirando: *es* el mejor de todos los mundos posibles... ¡Y no podría ser peor! Según afirman los filósofos, "deber" implica "poder", incluso en dominios que nada tienen que ver con el libre albedrío o la responsabilidad moral.

Aun cuando sea correcto abandonar nuestra fidelidad al principio "podría haber obrado de otra manera", en tanto prerrequisito para la acción responsable, subsiste todavía el problema de que en el determinismo (según los compatibilistas) nunca podemos hacer nada sino lo que de hecho hacemos. Slote señala que "esto se parece bastante a una recusación de nuestras creencias más caras y más profundamente arraigadas, de modo que vale la pena averiguar si los argumentos recientes pueden atacarse en *un punto anterior* a la conclusión de que todas las acciones son necesarias" (Slote 1982, pág. 9). Pero la recusación es más desagradable de lo que afirma Slote.

Si los incompatibilistas estuvieran en lo cierto repecto de nosotros, sería porque estarían en lo cierto respecto de todo: en el determinismo, *nada* puede hacer otra cosa que lo que en realidad hace. Por lo tanto, la conclusión debe ser la siguiente: puesto que en un mundo determinista, un átomo de oxígeno que nunca se combina con átomos de hidrógeno está determinado para no combinarse jamás con átomos de hidrógeno, es *físicamente* imposible que se combine con otros átomos de hidrógeno. ¿En qué sentido, entonces, sería verdadera la afirmación de que *puede* combinarse, como cualquier otro átomo de oxígeno, con dos átomos de hidrógeno?

Ayers denomina a esta endeble implicación del determinismo "actalismo" —sólo lo actual es posible (Ayers 1968, pág. 6)—. Algo anda mal, sin duda, en el actualismo; pero está tan equivocado que es harto improbable que su falsedad pueda ser discutida dentro de una *reductio ad absurdum* del determinismo. El argumento sería asombrosamente breve: este átomo de oxígeno tiene valencia 2; por lo tanto, puede unirse a dos átomos de hidrógeno para formar una molécula de agua (lo *puede* hacer ahora mismo, no importa si lo hace o no); luego el determinismo es falso. Existen impresionantes argumentos en la física que conducen a esta misma conclusión, pero éste no es uno de ellos.

Hume dice que todos deseamos que exista "una cierta soltura" en el mundo (*Tratado sobre la naturaleza humana,* II, III, 2). Esta soltura que impide que lo posible se reduzca a lo real, está presupuesta en el verbo "poder"; y como la necesitamos para muchas cosas, es importante conocer el significado del término "poder", no sólo para explicar la libertad o para las ciencias sociales, sino también para la biología, la ingeniería y, en realidad, para todo campo que se base sustancialmente en la estadística y en la teoría de la probabilidad.

¿A qué se refieren los biólogos, por ejemplo, cuando dicen que una característica determinada de una especie es *mejor* que alguna otra característica "posible"? Si hemos de describir coherentemente la tendencia adaptativa de la selección natural, tenemos que distinguir, con frecuencia, los diseños seleccionados como mejores (o como *no* mejores) que otros "posibles" diseños que la selección desechó.[9] Los biólogos nos aseguran que los unicornios no sólo no

[9] Véase, por ejemplo, el interesante artículo de Sober "The Evolution of Rationality" (Sober 1981, pág. 110), en el cual se habla de un "proceso de selección en el que muchas cualidades posibles simplemente no estén representadas". Nótese que tanto la negación como la aceptación del adaptacionismo presuponen la coherencia de los supuestos acerca de la posibilidad. (Véase Dennett 1983b, sobre los riesgos y beneficios de adaptacionismo.)

son reales: son imposibles, tan imposibles como las sirenas. (Supongo que esto se relaciona con la presencia de un cuerno único, en posición central, que viola el principio de bilateralidad.) Pero los biólogos también nos aseguran que hay muchas especies posibles todavía inexistentes y que es probable que nunca existan —caballos gordos, de patas cortas, que sólo sirven para comer o jirafas cuya piel carece de manchas—. Sólo un pequeño número de variaciones posibles aparece alguna vez.

En la teoría de la probabilidad se afirma que tirar la moneda tiene solamente dos resultados posibles: cara o cruz.

> Cuando se tira la moneda y X presencia el acontecimiento, X piensa que su hipótesis probablemente sea verdadera: la moneda caerá con el anverso hacia arriba (cara) o con el reverso hacia abajo (cruz). Caben, también, otras posibilidades: que la moneda caiga de canto. No obstante, aunque X dé por sentado el más crudo folclore de la física moderna, descartará, por imposible, el movimiento ascendente de la moneda hacia el espacio exterior rumbo, digamos, a Alfa de Centauro (Levi 1980, pág. 3).

Dondequiera miremos, nos encontramos con afirmaciones categóricas acerca de la clase de cosas que pueden hallarse en determinados estados, de los posibles resultados y de lo que es imposible pero no *lógicamente* imposible (autocontradictorio).

Este sentido elusivo del verbo "poder" parece no tener nada que ver con la *agencia*; sin embargo, se halla presente en ese terreno. En "Ifs and Cans", Austin (1961) critica el intento de definir "podría haber obrado de otra manera" como "habría obrado de otra manera si...", llenando los blancos con diferentes instancias. Numerosos filósofos (véase especialmente Chisholm, 1964a) criticaron hábilmente las objeciones de Austin. Pero más importante que esas objeciones y críticas que tanto interés suscitaron en el ambiente filosófico, es su áspera, inapelable y muy válida descalificación (en una nota al pie de página) del abordaje más promisorio al problema residual de la rana en el fondo del jarro de cerveza.

Austin señala, al pasar, que "hay algo plausible en la sugerencia de que 'yo puedo hacer X' significa que 'lograré hacer X si lo intento', y 'yo podría haber hecho X' significa que 'habría logrado hacer X, si lo hubiera intentado'"; pero en la famosa y extensa nota a pie de página desecha la sugerencia:

> Plausible, pero nada más. Supóngase que, jugando al golf, he errado el golpe y me maldigo porque podría haber embocado la pelota en el hoyo. Ello no significa que habría embocado la pelota si lo hubiera intentado, pues lo intenté y fracasé; tampoco significa que la habría

> embocado si las condiciones hubiesen sido diferentes; por cierto, podría haber sido el caso, pero me estoy refiriendo a las condiciones tal como eran exactamente y afirmando que podría haber embocado la pelota. De ahí mi fastidio. Asimismo, "puedo embocar la pelota otra vez" no significa que esta vez embocaré si lo intento, pues puede suceder que lo intente, fracase y aun así no me convenza de que no podría haberlo hecho. En realidad, un número mayor de intentos puede llegar a confirmar mi creencia de que en esa ocasión podría haberlo hecho, aunque no lo hiciera.
> Pero si luego de poner todo mi empeño, fracaso, *¿debe* de existir seguramente algo que causó mi fracaso y me hizo incapaz de tener éxito? En consecuencia, *no podría* haber embocado la pelota. Ahora bien, la moderna creencia en la ciencia y en su capacidad para explicarlo todo, puede inducirnos a aceptar el argumento. Pero tal creencia no es compatible con las creencias tradicionales que el verbo "poder" encierra. Según estas creencias, la habilidad, el poder o la capacidad humanas no están intrínsecamente sujetos a producir buenos resultados, y no hay ninguna razón para ello (¿o acaso la mala suerte o el no hallarse en buen estado cuentan, a veces, como razones?).

Pero entonces, ¿qué debería desecharse, según Austin: "la moderna creencia en la ciencia' o 'las creencias tradicionales que el verbo 'poder' encierra?". Austin no lo dice y nos deja en un callejón sin salida; pero se trata de un *impasse* ilusorio. La ciencia moderna necesita el mismo verbo "poder" que necesitan las creencias tradicionales acerca del obrar humano. Y lo que se debe desechar es la insistencia de Austin cuando afirma que "me estoy refiriendo a las condiciones tal como eran exactamente". Como vimos, referirse a condiciones de precisión microscópica no reditía ninguna ganancia; cuando se habla de lo que alguien (o algo) *puede hacer*, lo que interesa es siempre algo general.

Honoré (1964) trató lúcidamente el punto en un comentario crítico, raras veces mecionado, sobre el trabajo de Austin. Según este autor, cabe distinguir dos sentidos del verbo poder: un sentido particular y un sentido general. El sentido particular es casi degenerado "equivale prácticamente a 'querer' y tiene su misma fuerza de predicción" (pág. 464). Usado en pretérito, el verbo poder (particular) sólo es apto para describir logros: "de esta manera, 'pude verlo a usted en la maleza' se usa correctamente sólo cuando yo he logrado verlo a usted en la maleza".

> El éxito o fracaso, suponiendo que se ha hecho o se hará un esfuerzo, es el factor que determina el uso de la noción: si el agente lo intentó y fracasó, no pudo realizar la acción; si lo intentó y lo logró, pudo realizarla. (Honoré 1964, pág. 464.)

El sentido general del verbo poder es más útil; en el caso del agente, asigna habilidad o destreza, y en el caso de las cosas inanimadas, la clase de potencialidad que discutimos en el capítulo 5 (por ejemplo, los diferentes estados en los que puede hallarse una cosa). Como vimos, este sentido general es una noción claramente *epistémica*; es decir, está generada por la necesidad, por parte de quienes planean y se controlan a sí mismos, de dividir el mundo en cosas y en sus "estados" que son posibles *según lo que saben.*

La tradición filosófica considera distintos tipos de posibilidad, entre ellas:

(a) La posibilidad *lógica* o alética: el complemento de la imposibilidad lógica; algo es lógicamente posible cuando se lo puede describir en forma consistente; es lógicamente posible que haya unicornios en el jardín pero (si los biólogos están en lo cierto) no lo es biológica o físicamente.

(b) la posibilidad *física* o nómica: algo es físicamente posible cuando no viola las leyes de la naturaleza o de la física (*nomos* significa ley en griego). Es físicamente imposible viajar a una velocidad mayor que la de la luz, aunque se pueda describir tal proeza sin incurrir en contradicción.

(c) La posibilidad *epistémica*: algo es epistémicamente posible para *Jones* cuando es compatible con todo lo que Jones ya *conoce*. Se considera, generalmente, que la posibilidad epistémica es subjetiva y relativa, a diferencia de las posibilidades lógica y física, que se suponen totalmente objetivas.[10]

En las discusiones filosóficas sobre el libre albedrío se acostumbra, una vez que se la ha individualizado, descartar la posibilidad epistémica por considerarla de escaso interés en el contexto.[11] La descalificación de Austin es una de las más breves. Según dicho autor, además de los dos sentidos, ya analizados, de "podría haber", existe un tercero, según el cual "podría haber obrado de manera diferente" significa "me sería posible, por todo lo que ya se sabe, haber obrado de manera diferente". En mi opinión, se trata de un vulgarismo en el que "podría" se usa incorrectamente en lugar de "me sería posible". De todos modos, el punto no es relevante aquí (Austin 1961, pág. 207).

[10] Véase Hacking 1975, sobre las complicaciones y calificaciones importantes.

[11] Véase, por ejemplo, el breve parágrafo de Van Inwagen (1983, págs. 9-10) y el abordaje mucho más cauteloso de Ayers, que comienza de la siguiente manera: "Las discusiones sobre el poder y la potencialidad, especialmente cuando se suscitan en el ámbito del libre albedrío, giran obsesivamente en torno de la noción de posibilidad epistémica y de las nociones afines de certeza e incertidumbre, de lo predecible e impredecible. Ello no es una mera confusión..." (Ayers 1968, págs. 13 y siguientes).

Es una vergüenza que los filósofos no se hayan interesado en una cuestión que constituye una de las claves para resolver el enigma de "poder". La idea útil que encierra el verbo, la idea en la que se basan no sólo los proyectos y deliberaciones, sino también la ciencia, es un concepto de posibilidad —y, a partir de él, naturalmente, los conceptos recíprocamente definidos de imposibilidad y necesidad— que son esencialmente "epistémicos", aunque no lo parezcan a primera vista.

Como Slote señala en su innovador artículo "Selective Necessity and Free Will" (1982), la clase de conceptos de posibilidad y de necesidad que se usan en esos contextos obedecen a principios modales diferentes del concepto de necesidad alética "clásica". En particular, tal necesidad no es "aglomerativa", esto significa, según Slote, que es cerrada respecto de la introducción de la conjunción.[12] El autor ilustra el concepto describiendo un encuentro "accidental". Jules se encuentra en el banco con su amigo Jim y piensa que el hecho es un accidente feliz (y en verdad lo es). Pero que Jules esté allí no es un accidente, pues, por razones de trabajo, concurre al banco todos los miércoles por la mañana. Tampoco es un accidente que Jim se encuentre allí, puesto que fue enviado al banco por su jefe. Luego, que Jules se encuentre en L en el tiempo t no es un accidente y tampoco lo es que Jim se encuentre en L en el tiempo t. Pero que Jules se encuentre en L en el tiempo t *y* que Jim se encuentre en L en el tiempo t, *eso sí* es un accidente (Slote 1982, en especial las págs. 15-17).

Desde un punto de vista limitado, se trataría, aparentemente, de un accidente o de una coincidencia. Pensamos que si supiéramos mucho más de lo que saben Jules y Jim juntos, habríamos podido predecir la convergencia de Jules y Jim en el banco. *Para nosotros,* su encuentro no habría sido "accidental". Sin embargo, éste es, justamente, el concepto de accidentalidad que se necesita para describir "la independencia" de los poderes o habilidades de una cosa con respecto a las condiciones iniciales o a las condiciones del contexto en las cuales se ejercen esos poderes. No es accidental, por ejemplo, que este insecto en particular tenga un modelo de vuelo evasivo (pues la evolución lo diseñó a tal efecto). Y no es accidental que el ave de presa que atrapa esos insectos posea los genes que

[12] Slote subraya la posibilidad de que la forma de la "necesidad selectiva" que describe sea, en realidad, una forma encubiertamente epistémica. No obstante, nos ofrece un sinnúmero de observaciones que nos conducen en esa dirección. Schotch y Jennings (1980) fundamentan filosóficamente las dudas sobre el atractivo universal de los principios de plena agregación o aglomeración en la lógica modal, pero fracasan cuando analizan los casos (más apremiantes) que Slote presenta.

posee (pues también fue diseñada para ello). Pero es un accidente —feliz para el ave y su progenie, desafortunado para el insecto— que un pájaro con ese tipo de genes atrape justamente a este insecto con modelo de vuelo evasivo. Y, gracias a los accidentes felices de esta índole, se llega a diseñar mejores pájaros —y mejores insectos—; gracias a la conjunción de millones de accidentes felices, aparece el espacio de "posibilidad" donde transcurre la evolución. El eminente biólogo Jacques Monod se refiere al importante papel que desempeña la suerte en la evolución e ilustra lo que denomina "la coincidencia absoluta" (Monod 1972, págs. 112 y sigs.) con un ejemplo asombrosamente parecido al de Slote:

> El doctor Brown se encamina al domicilio de un nuevo paciente para hacer una visita de emergencia. Mientras tanto, Jones, el hombre que pertenece a una empresa de reparaciones, está haciendo una reparación de emergencia en el techo de un edificio vecino. Cuando el doctor Brown pasa por el edificio, Jones, inadvertidamente, deja escapar el martillo, cuya trayectoria (determinista) intercepta la trayectoria del médico, que muere por fractura de cráneo. Se dice de Brown que fue víctima del azar (pág. 114).

Pero cuando Monod define las condiciones en las que ocurren tales coincidencias cae, aparentemente, en la trampa del actualismo. Para que haya revolución tiene que haber accidentes, dice Monod, y los accidentes pueden ocurrir, "a menos que volvamos al mundo de Laplace, en donde la suerte está excluida por definición y donde el doctor Brown está condenado a morir a causa dle martillo de Jones desde el principio de los tiempos" (pág. 115).

Si "el mundo de Laplace" significa un mundo determinista, Monod está equivocado. La selección natural no necesita una "coincidencia absoluta" sino una independencia práctica" —la clase de independencia que se mostrón en el caso de Jules, Jim, Jones y el doctor Brown: cada uno seguía su propio camino pero "ocurrió" que se interceptaron, como los naipes de un mazo que se baraja de manera determinista y "ocurre" que van saliendo en un orden consecutivo—. ¿Sería posible la evolución de un mundo determinista, laplaceano, donde la mutación fuera causada por un proceso no aleatorio? Sí, pues lo que la evolución exige es un generador de materia prima que no tenga configuraciones, no un generador de materia prima que no tenga causas. Es posible que los efectos en el nivel cuántico desempeñen un papel verdaderamente importante en producir mutaciones, pero no es un papel que la teoría requiera.[13]

[13] La interesante discusión de Monod (véanse especialmente las págs. 77-80 y 111-

No es evidente que un proceso requiera (o pueda detectar) el azar "genuino" u "objetivo", sea de tipo mecánico—cuántico, sea de tipo matemático, incomprensible respecto de la información. (Chaitán, 1976, presenta una demostración gödeliana de que no existe ningún procedimiento de decisión que determine si una serie es matemáticamente aleatoria.) Incluso en matemática, donde el concepto de azar objetivo tiene aplicación crítica en las demostraciones, existen casos de los que podría llamarse "indiscernibilidad" práctica.

En la aritmética, el teorema de Fermat-Lagrange afirma que todo número natural es la suma de cuatro cuadrados perfectos:

$$n = x^2 + y^2 + z^2 + w^2 .$$

Supongo que la demostración del teorema es bastante simple, pero encontrar los valores de x, y, z y w para un n dado puede resultar muy tedioso. Existe un algoritmo directo, "brutal", que siempre encontrará estos valores utilizando el simple y exhaustivo método de ensayo y error, pero posee un grave inconveniente: necesita, como término medio, 2^n pasos para terminar la operación; es decir, que para un número pequeño como el 203, el algoritmo no podrá hallar la respuesta antes de que se haya extinguido todo calor en el universo. Evidentemente, no es lo que en la jerga se llama un *algoritmo factible*, aunque *en principio* (como dirían los filósofos) siempre provee la respuesta correcta.

Pero no todo está perdido. Rabin y otros desarrollaron los llamados algoritmos aleatorios que se basan, de manera extremadamente antiintuitiva, en la generación del número aleatorio. Rabin descubrió uno de esos algoritmos para hallar valores para el teorema de Fermat-Lagrange. No hay garantía lógica de que se encontrará la respuesta correcta en un tiempo menor que el demandado por el algoritmo que usa "la fuerza bruta", pero el plazo esperado es sólo de $(\log n)^3$ pasos, un número manejable y pequeño incluso para grandes valores de n. La probabilidad de que el plazo de tiempo sea mucho más largo o mucho más corto es prácticamente insignificante. La demostración formal de que éste es el plano esperado hace mención especial de la invocación de las secuencias aleatorias en el algoritmo.

En el mundo actual de los ordenadores de *hardware* ¿es importante si el ordenador utiliza secuencias aleatorias genuinas o secuencias seudoaleatorias? Esto es, si escribiéramos el programa

---

117) es ambigua y problemática en ese punto. Véase también la valiosa discusión sobre la pregunta en Winsatt (1980, especialmente sección 3), "Periodic, Almost-Periodic and Chaotic Behaviour in Simple Models of Population Growth and Regulation".

de Rabin y lo ejecutáramos en un ordenador que no tiene un generador de radio del número aleatorio sino que se basa en un algoritmo generador del número seudoaleatorio, ¿funcionaría este método abreviado y más barato, o bien los intentos por encontrar los valores para un *n* particular demandarían más tiempo que el número esperado de pasos en virtud del carácter no aleatorio, oculto e indetectable, de la secuencia? ¿El sistema de números podría castigar al matemático por sondear sus secretos mediante una simple exploración seudoaleatoria? Según lo indica la experiencia acumulada hasta el momento, uno puede arreglárselas muy bien con las secuencias seudoaleatorias.[14]

Pero, seguramente, la mera "indiscernibilidad" práctica, aun en el límite, no es la posibilidad objetiva, real. Esta es la intuición que debemos examinar ahora. Se halla en la base misma del rechazo que experimentan los filósofos por la posibilidad epistémica en tanto piedra fundamental del libre albedrío. La llamada física "clásica" o newtoniana es determinista pero, como han señalado recientemente varios físicos, nadie que no fuera un demonio laplaceano *infinito* podría predecir —*con los principios de Newton*— muchos de los fenómenos macroscópicos más triviales en un mundo newtoniano, pues se necesitaría una precisión infinita en la observación inicial. Esto es, los errores de observación, aunque fueran minúsculos, se propagarían y crecerían exponencialmente (Berry 1983 y Ford 1983).

En la física newtoniana existen sistemas estables (algunos pocos valiosos) y sistemas caóticos o inestables. "En los sistemas no caóticos el error se propaga menos rápidamente... incluso un hecho notorio del pasado basta para determinar con precisión un hecho notorio del futuro." Los eclipses, por ejemplo, se pueden predecir con siglos de anticipación. Pero "una órbita caótica es aleatoria e incalculable; su contenido de información es infinito e incomprensible a la vez" (Ford 1983, pág. 7). La trayectoria de la bola de un juego electromecánico (el ejemplo es de Berry), luego de rebotar en veinte lugares distintos en unos pocos segundos, es impredecible *en el límite*, superando los límites de precisión de cualquier dispositivo de observación imaginable. Ahora bien, éste es un resultado "meramente epistémico". ¿Cuál es su relación con el libre albedrío?

Pienso que es la siguiente: los sistemas caóticos son la fuente de la independencia *práctica* (cabría decir: infinitamente práctica) de las cosas que desordena el mundo y lo transforma en un sitio de

[14] Dicho personalmente por Rubin en una reunión en la Academia de Ciencias de New York, abril de 1983. Véanse también Rabin 1980 y Jauch 1973 (discutido en Hofstadter 1979, págs. 408-409).

continuas oportunidades. Estas oportunidades no son sólo *nuestras* sino también de la Madre Naturaleza (y de los átomos de oxígeno que en ocasiones unen sus fuerzas con los átomos de hidrógeno). Cuando decimos que el mundo se divide en sistemas estables y predecibles y en sistemas caóticos e inestables, no nos estamos refiriendo exclusivamente a *nuestras* limitaciones epistémicas, sino al mundo mismo, pues es una afirmación acerca de la capacidad del mundo de ser predecible por cualquier sistema de predicción, no importa cuán poderoso sea éste. No hay una perspectiva más alta (a menos que consideremos la perspectiva de un ser infinito) desde la cual las colisiones "accidentales" de trayectorias localmente predecibles sean en sí mismas predecibles y, por tanto, no sean "accidentes".

Este contraste entre lo estable y lo caótico es lo que fundamenta nuestra división del mundo en elementos notorios y constantes y en elementos que *deben* ser tratados estadística o probabilísticamente (es decir, promediándolos y borrando así sus contornos, o bien tratándolos como miembros equiposibles de algún conjunto de alternativas). Y esta división del mundo no es sólo nuestra, sino también de la Madre Naturaleza. Dado que, según lo que sabe (o podría saber), es posible que estos insectos se crucen (alguna vez, en algún lugar) con estos pájaros insectívoros, sería mejor que estuvieran diseñados con algún mecanismo de evitación. Esto le da un cierto poder que (en general) les será beneficioso.

(Estas observaciones demasiado esquemáticas sobre el verbo "poder" son, en el mejor de los casos, un gesto que señala la superficie final, acabada, de esta parte de mi estatua del agente libre. Se trata de un área donde es necesario trabajar mucho más y parte de ese trabajo está, seguramente, más allá de mis posibilidades. Pero si este primer y somero relevamiento de la zona no ha sido desacertado, la tarea que aún queda por realizar al menos apartará la investigación de las superficies anticuadas y excesivamente trabajadas hacia nuevos espacios.)

7

La necesidad del libre albedrío

Fue entonces en esta esfera,
la esfera de las obligaciones legales,
donde se originó el mundo de los conceptos
morales de "culpa", "conciencia", "deber",
"sacralidad del deber". Sus comienzos
fueron como el comienzo de toda grandeza
en la Tierra, particularmente sangrientos
y por largo tiempo.
—Nietzsche, *La genealogía de la moral.*

1. La refutación del nihilismo

Las variedades del libre albedrío que vale la pena desear son aquellas que garantizan la dignidad y la responsabilidad humanas. Ciertas concepciones filosóficas fomentan, sin embargo, el anhelo por la "libertad contracausal" o por la "causalidad del agente", convenciéndonos, con razón o sin razón, de que tal bendición metafísica es la condición necesaria para el tipo de libre albedrío que debe poseer todo ser humano moral, digno y responsable. Si resulta que no existe ese tipo de libertad, la idea misma de responsabilidad pierde su fundamento.

Es interesante notar que las personas, cuando tropiezan con esta idea, no la toma generalmente como la perspectiva de unas felices vacaciones, es decir, como la posibilidad de hacer lo que les place sin correr el riesgo de ser o de sentirse culpables. Y nos sorprende, pues si la responsabilidad fuera un concepto sin fundamento, el hecho debería ser liberador, jubiloso, digamos, un hálito fresco barriendo los oscuros dogmas del pecado y la culpa. No obstante, las personas se empeñan en ser consideradas responsables. ¿Por qué?

Porque la perspectiva de no ser responsables amenaza nuestra dignidad, y nuestra dignidad nos importa. Quizás estamos en un error, fascinados por una visión del mundo anacrónica y por una moral que ya no es defendible en la lúcida era de la ciencia (Skinner 1971).

O quizá nuestra lealtad obedezca a razones más oscuras. ¿Cómo interpretar, si no, el deseo de que los *otros* también sean responsables? No queremos que Hitler, Nixon o el muchacho que nos roba las tapas de las ruedas queden impunes. Tal vez pensamos que si el precio a pagar por la responsabilidad de los otros es la afirmación de nuestra propia responsabilidad, vale la pena cerrar el trato. Pero ¿por qué desearemos con tanta vehemencia que los demás sean considerados responsables? ¿Podría tratarse de un rasgo inherente a nuestra persona, un rasgo vengativo racionalizado y presentado bajo una apariencia civilizada gracias al barniz de una doctrina moral?

Existen, no cabe duda, elementos atávicos y sanguinarios en la visión ordinaria de la responsabilidad, de la dignidad y de la culpa. Hay un "cuento" muy brillante sobre el tema en el libro de Nietzsche *La genealogía de la moral*. Nietzsche afirma que los conceptos morales de culpa y castigo se originaron en los conceptos *legales* de débito y crédito: según éstos, "el acreedor podrá infligir toda clase de iniquidad y torturas en el cuerpo del deudor" (Nietzsche 1887). Entre las deudas reconocidas por nuestros predecesores estaban las deudas de gratitud con los antepasados. Y como el sentimiento de estar en deuda se acrecentó (y la deuda no se podía pagar), los antepasados acreedores se transformaron en un Dios acreedor y la deuda, en una carga intolerable. El"cuento" sienta las bases para la osada especulación de Nietzsche, en la que describe la fundamentación racional (sin duda "flotante") que subyace a la invención inconsciente de la culpa consciente hecha por los sacerdotes del *resentimiento*:

> De pronto nos encontramos ante el paradójico y horripilante recurso que deparó transitorio alivio a la humanidad atormentada, ese golpe de genio del cristianismo: Dios mismo se sacrifica a sí mismo por las culpas de la humanidad, Dios mismo se paga a sí mismo, Dios, el único ser que puede redimir lo que es irredimible para el hombre mismo el acreedor se sacrifica a sí mismo por su deudor, por *amor* (¿puede creerse esto?), por amor a su deudor! (pág. 92).

El "cuento" es demasiado para estar totalmente equivocado; pero aun en el caso de ser verdadero, el relato (o cualquier otro sobre un tema similar) no demuestra que no pueda ser también una venganza de nuestro esquema conceptual, o al menos de una parte

importante de ese esquema. Las verdaderas conclusiones nacen, en ocasiones, de razonamientos falaces y más de una gran idea ha surgido de la actividad inconsciente.[1]

Pienso que el deseo de preservar "nuestro mundo de conceptos morales" es válido y cabe defenderlo. La distinción que se establece entre los agentes morales responsables y los seres cuya responsabilidad está disminuida o nula, es una distinción importante, real y coherente; es coherente incluso en aquellos casos donde su aplicación es problemática; traza una línea divisoria empíricamente real, puesto que no todos caemos del mismo lado y, sobre todo, la distinción nos interesa: el uso que hagamos de ella desempeñará un papel crítico en la calidad y el significado de nuestra vida.

Mi defensa se apoyará en un argumento que demuestre que *es racional* valorar el libre albedrío y desear la propia responsabilidad: no cabe otro argumento para defender tales conceptos. Sin embargo, se lo tomará inevitablemente como una petición de principio, puesto que se apela, en apariencia, al juicio racional de los agentes circunstantes. ¿Acaso no doy por sentado que nosotros —los lectores y el autor— somos precisamente la clase de agentes responsables que mi argumento pretende defender? Además, presupone que *algunas cosas importan*; es decir, que algunas cosas son para bien y otras para mal. Sin este supuesto, nuestra pretendida racionalidad no tendría en dónde aferrarse. Pero esta hipótesis antinihilista ¿no es una petición de principios todavía más obvia?

A lo largo de este libro, jamás puse en duda la racionalidad de los lectores o la mía propia; tampoco cuestioné la significación de la empresa en la que me he comprometido. La misma decisión de escribir un libro —un libro tradicional, compuesto de oraciones y no, digamos, de una colección de páginas encuadernadas y untadas con diversos dulces y jaleas— puede parecer una petición de principio en favor del libre albedrío. Si éste es el caso, entonces quienes han escrito artículos y libros negando la realidad del libre albedrío se hallan en una situación aún más embarazosa: tienen que aconsejar al lector (o al menos simular que lo hacen) que aconsejar es inútil.[2] Pero existen otras alternativas tales como el silencio o el suicidio,

[1] Nietzsche mismo llamó la atención sobre este punto (ahora conocido) de la teoría de la evolución: la *raison d'être* de un órgano, conducta o práctica puede cambiar bajo la presión de las circunstancias. "Las causas del origen de una cosa, de su eventual utilidad, de su actual empleo y lugar en el sistema de las finalidades se encuentran a mundos de distancia."

[2] A menudo se ha declarado que los intentos de Skinner (1971) por convencer a los lectores de que no eran agentes libres y dignos son sistemáticamente autoembrutecedores. Para otro debate sobre el tema, véase Wolf (1981, especialmente pág. 395).

y aun puede subsistir la fuerte sospecha de que hemos estado apartando los ojos de visiones más profundas y terroríficas.

Examinemos esta sospecha. He dado por supuesta la significación de mi empresa, pero bien puedo equivocarme, pues podría ser tan defectuosa que careciera de sentido. La concepción que niega el sentido de *toda* empresa es el *nihilismo*, que sostiene que *nada* tiene importancia. Muchos sospechan que si el determinismo es verdadero, también lo es el nihilismo. Luego, para quienes albergan esa sospecha, mi proyecto será, en el mejor de los casos, insuficiente, puesto que no he tomado en cuenta el nihilismo. Más allá del reconcimiento de mi negligencia ¿qué puedo decir respecto de esa teoría?

No puedo decir que, si el nihilismo resultara verdadero, sería una verdadera lástima, una vergüenza flagrante o un ejemplo más de la "injusticia cósmica". En efecto, si el nihilismo fuera verdadro, *todos* los juicios de valor serían ilusorios. Hechos concretos tales como el dolor o la desgracia de las personas no significarían nada y lamentarnos de los problemas sería tan desatinado como lamentarnos de que la raíz cuadrada de dos no sea uno y medio. El nihilismo puede ser verdadero, de acuerdo; puede, incluso, inferirse del determinismo, como se afirma (aunque no veo la razón). No obstante, podemos suponer que es falso. ¿Cómo puedo hacer una afirmación tan soberbia y presuntuosa sin tomarme siquiera el trabajo de considerar con algún detalle su posibilidad? Ahora bien, si el nihilismo es verdadero, y debemos pensar seriamente esa posibilidad, entonces resultará falso (puesto que si *debemos* hacer *algo* el nihilismo es falso). Luego, cabe también suponer que es falso, ya que el hecho de que sea o no falso no tiene la menor importancia y podemos hacer lo que nos plazca. El nihilismo, literalmente hablando, es una posición *descartable*.

Y como puede que no haya nada mejor que hacer, afirmaremos, entretanto, la confianza en nuestro punto de partida. Podemos suponer que existen cosas que importan y que somos suficientemente racionales, de modo que poseemos un punto de apoyo para el intento de comprender el mundo. No veo cómo se podría criticar con coherencia estos supuestos (lo que no equivale a decir que sean verdaderos, sino que todavía quedan posibilidades menos radicales, que vale la pena analizar). Una cosa es suponer que existe lo bueno y lo malo, lo mejor y lo peor, un espacio para la esperanza y el arrepentimiento, y otra muy distinta dar por sentado que los conceptos morales tradicionales son correctos. Es probable, por ejemplo, que el concepto de responsabilidad valorado por la moral tradicional (occidental) sea sutilmente inconsistente y que se imponga una revisión e incluso un rechazo de dicho concepto y de sus afines: la

culpa, el mérito, el encomio moral, el castigo, para mencionar los más importantes. Quizá tengamos que descender de nuestro status tradicionalmente elevado de agentes morales para garantizar una moral defendible.

2. La responsabilidad disminuida y el espectro de la exculpación creciente

Vimos en el capítulo 2 que, si para ser un auténtico agente moral se necesitara poseer una perfecta voluntad kantiana, infinitamente sensible a la dulce voz de la razón, podría no haber agentes morales en el universo. También vimos, en el capítulo 4, la imposibilidad de que un yo se construya enteramente a sí mismo y sea ciento por ciento responsable de su carácter. Pero hemos visto que la perspectiva es halagüeña: nosotros, seres finitos e imperfectos, somos aproximaciones bastante fidedignas de esos absolutos imaginarios. Debemos buscar una perspectiva similar para el concepto de responsabilidad moral.

Aunque deseamos afirmar que todos los seres humanos son responsables, a menudo juzgamos intuitivamente que un determinado individuo tiene la"responsabilidad disminuida", sea por enfermedad, sea por circunstancias particularmente penosas durante la crianza o en el momento de la acción. También juzgamos que no son responsables los animales no humanos, los infantes y los mentalmente incapaces. Mas luego de haber excusado a todos (y por buensa razones) cabe preguntarse si quedarán responsables, dado que todos somos imperfectos en alguna medida.

En nuestros mejores momentos nos comportamos "responsablemente"; es decir, moralmente. Pero ¿cómo deberíamos tratar los casos en los que las personas muestran un diseño racional muy inferior al de sus mejores momentos? ¿Somos igualmente responsables de nuestras buenas y malas acciones? Según Kant, existe una asimetría: sólo somos responsables de lo que hacemos bien. (Para una versión contemporánea de la misma asimetría, véase Wolf, 1980). Sócrates fue el primero en plantear el tema en su afirmación "paradójica" de que nadie desea hacer el mal a sabiendas. Desde entonces, se ha convertido en un tema perenne del debate filosófico, aunque se lo haya planteado de una manera indirecta y caprichosa. ¿De qué tenemos miedo? Tememos, ciertamente, que nadie merezca el castigo que la sociedad le inflige, ya que todos los malhechores *ipso facto* se engañan, padecen desórdenes mentales o son, en cierto sentido, radicalmente ignorantes.

Es necesario, pues, abocarnos a la tarea de trazar (y defender) una línea divisoria entre la patología que exculpa (la que se alega

en "casos de insania" evidentes) y los diversos tipos de deficiencia que no impiden la auténtica culpabilidad de los agentes; es decir, buscamos una cierta libertad de acción para nosotros, los pecadores en medio de los santos y de los monstruos.[3] El problema no incumbe sólo a la justicia. Dependen de ello nuestras creencias morales más arragaidas, y no cabe duda de que nuestras intuiciones están en conflicto. Por ejemplo, despreciamos a Nixon y a la vez nos maravillamos de que se engañara de un modo tan ingenuo y grotesco. Luego, o bien Nixon es en realidad la víctima de ese engaño (¡pobre muchacho!) o bien es doblemente culpable por haber perpetrado el (auto)engaño, además de los otros delitos. El autoengaño, dejando de lado si se trata o no de una enfermedad conocida, es esencialmente ubicuo. Si puede eximir de culpas, ¿nos eximirá a todos? ¿Siempre?

Sentimos la tentación de aferrarnos a alguna versión de la responsabilidad cósmica como baluarte de la "exculpación que paraliza", pero incurrir en absolutismos filosóficos de esa índole sería un grave desacierto.

Si vamos a afirmar nuestra responsabilidad, se tratará de una responsabilidad modesta, naturalizada, ligeramente disminuida, ya que no somos ángeles. Comentando el fracaso del absolutismo de Kant, Williams señala que "ninguna característica humana relevante para la valoración moral puede dejar de ser únicamente empírica, sujeta a condiciones empíricas, a la historia psicológica y a las variaciones del individuo" (Williams 1973, pág. 228).

¿En qué casos somos responsables de nuestras malas acciones? Para responder a la pregunta consideraremos, en primer término, la finalidad de afirmar que las personas son responsables. Todo juicio de responsabilidad, sea en la corte, sea en privado, sienta las bases para una decisión sobre sanciones: ¿qué haremos?, ¿qué respuesta cabe dar al agente responsable?

> La evaluación moral es importante porque incide en el comportamiento *vis à vis* de las personas evaluadas. Debemos encarcelar, amenazar o culpar a los delincuentes, no a su comportamiento; y nuestra conducta es racional sólo en tanto y en cuanto sea compatible con lo que hemos descubierto acerca de los delincuentes a través de su comportamiento. (Cummins 1980, pág. 212.)

Podemos tomar la decisión de perdonar, reprochar, denunciar, increpar o incluso castigar al reo. Cada una de estas alternativas,

[3] Slote me ha recordado la teoría medieval según la cual los más santos y bienaventurados son aquellos incapaces de pecado (*"non posse pecare"*). Véase, por ejemplo, San Agustín, *De libero arbitrio*; San Anselmo, *De libero arbitrio* (cap. 1, col. 489-491); Santo Tomás, *De Veritate*.

que ocupa un sitio diferente y posee un fundamento racional diferente, contribuye a afianzar la creencia en la responsabilidad del prójimo. Pero, a fin de simplificar las cosas, me limitaré a la respuesta más explícita (pública, codificada e instituida) de todas: el castigo. Existe una idea que podríamos denominar estratégica inserta en el fundamento racional del castigo, que, en mi opinión, constituye la clave para entender nuestra posición como agentes morales responsables. La analizaré presentando una "reconstrucción racional" simplificada de lo que considero el fundamento racional (hasta cierto punto "flotante") de la institución de la ley penal con su elaborado sistema de represión y castigo. Puesto que mi propósito es el de exponer una idea y no una defensa irrecusable de la institución de la ley penal, pasaré por alto las sutilezas.[4]

¿Por qué deseamos castigar a la gente que comete delitos? Cabe especular con Nietzsche sobre las oscuras corrientes de violencia subterránea que perturban nuestra psiquis, pero cualesquiera sean los méritos de tales hipótesis hay buenas razones que justifican la existencia de esta institución particularmente humana. Podemos individualizar con rapidez la clase de males que nos gustaría reducir a un mínimo; tenemos motivos para suponer que si se prohíben las causas y se refuerza la prohibición mediante sanciones, disminuirá seguramente la frecuencia de esos males en nuestra sociedad. En primer término, se desprende del concepto mismo de racionalidad que si los miembros de la sociedad son (a veces) racionales, comprenderán que no conviene a sus intereses que los atrapen cometiendo actos delictivos. Y tenemos suficiente evidencia empírica de que la ciudadanía (considerada como un todo) responde a esas instituciones.[5] Las leyes (respaldadas por la sanción) establecen una diferencia en la dirección deseada, aunque reconocemos que esos efectos no son ideales.

En un mundo ideal, las personas disciernen cuanto es correcto hacer y lo hacen sólo por esa razón. No se necesitan leyes ni tampoco un sistema de sanciones. Todos se comportan como ángeles. En una palabra, es el cielo en la tierra. En un mundo un poco menos ideal (digamos "un peldaño más abajo") necesitaremos un sistema de leyes a causa de la agresividad y el egoísmo de las personas (si son como nosotros). Pero el sistema será perfectamente disuasorio porque las personas serán muy racionales (a diferencia de nosotros). Todo

[4] Un clásico examen de estas cuestiones se encuentra en Hart 1968. Véase especialmente su ensayo: "Prolegomenon to the Principles of Punishment", págs. 1-27.

[5] No debería subestimarse la dificultad de establecer (en la corte de la ciencia) hipótesis particulares acerca de los efectos disuasorios de leyes particulares. Véase Gibbs, 1975.

el mundo considerará obvio, tan obvio como tener una nariz en el medio de la cara, que el delito no produce beneficios y, por lo tanto, no se cometerán delitos. Los jueces, la policía y los carceleros serán nombrados y adiestrados y se quedarán sentados, como la guardia suiza, esperando la orden de entrar en acción, pero dudando de que llegue alguna vez en sus vidas.

¿Por qué no nos hallamos en una situación similar? Si somos en realidad el *homo sapiens*, el "animal racional", ¿por qué nuestras prisiones están atestadas y nuestros jueces sobrecargados de trabajo? Porque, quizá, las fuerzas de represión no son suficientes y, entonces, las personas, siendo efectivamente racionales, consideran que en ciertas ocasiones el crimen *paga*, o es probable que pague, de modo que vale la pena correr el riesgo. El poder disuasorio de las leyes es (idealmente) una función de la percepción de un hecho verosímil: la captura y la severidad de la pena a infligirse. Cuando uno de estos factores aumenta, aumenta la disuasión. Las infracciones de tránsito (por ejemplo, no detenerse ante la luz roja) podrían reducirse a un mínimo instalando un patrullero en cada bocacalle las 24 horas del día (en ese caso, la captura sería harto probable); o bien condenar a los infractores a cadena perpetua. Pero como la ciudadanía es racional, decide que cada una de las opciones —o una combinación de ambas más un poco de agua— no vale lo que cuesta. Es preferible no respetar unos pocos semáforos que invertir una buena parte de nuestra riqueza en las fuerzas de represión; o en inflingir castigos tan brutales (y embrutecedores) a los delincuentes. De manera que escatimar nuestro dinero es (aparentemente) racional.[6]

La institución óptima será la que"tolere" un cierto monto de delito, captura y condena, puesto que una inversión mayor en las fuerzas que respaldan la ley se traducirá en una rápida merma de los ingresos. Esto no significa que el sistema actual no necesite profundas reformas, sino que sería irracional aferrarse a la esperanza de un sistema disuasorio *perfecto*. En consecuencia, los que infringen la ley siempre estarán entre nosotros y el carcelero no será jamás una figura meramente decorativa.

Pero aun reconociendo el valor de reducir a un mínimo los delitos (tratando, al mismo tiempo, de reducir los costos que insumen la represión y el castigo), se impone también una reforma de las

[6] Pero esperen un momento: ¿sería tan costoso lograr que las penas fueran lo bastante severas y creíbles como para disuadir a todos? Sí, y por la consiguiente razón: las prisiones estarían prácticamente vacías, pero las fuerzas policiales tendrían que ser muy numerosas, pues surgiría la tentación de evadirse y de resistir el arresto, a causa de la severidad de las penas. Y desde luego, existen otras objeciones muy válidas contra la propuesta, aunque se trate de una reforma en permanente consideración.

leyes para que se adapten mejor a las circunstancias. La disuasión depende de varios factores: uno de ellos es la "publicidad". La disuasión solamente logra su propósito con las personas que *conocen* la ley y *comprenden* sus requisitos y sanciones. Hay individuos que no reúnen estas condiciones y cometen delitos porque el poder disuasorio de la ley no los alcanza. Luego, una parte del coste de la institución de la justicia corresponderá a la educación pública: las leyes secretas no disuaden.

Combinar un programa vigoroso de educación pública con un artificio legal en cierto modo perentorio y enérgico, es una manera eficiente de lograr un nivel satisfactorio de conocimiento (compatible con el coste). Ambos elementos deben obrar en conjunto: ignorar la ley no es excusa.

Esta última condición permite mantener una moderada curiosidad hacia las leyes y sus reformas; pero tal pretensión sería ofensiva y racionalmente inadmisible en el caso de faltar un programa de información pública "suficientemente" enérgico. Puesto que el Estado hace su parte al facilitar información, le cabe a la ciudadanía la responsabilidad de conocerla, lo que no es pedir mucho. Es importante advertir que el principio de que la ignorancia no es excusa, tiene su dosis de arbitrariedad: no supone que no *podría haber* fundamentos intuitivamente válidos para alegar ignorancia sino que declara que tales alegatos *no serán (normalmente) ni siquiera tomados en cuenta*. La vida es corta; la ley debe ser eficaz y es necesario trazar la línea en alguna parte.[7] (Luego veremos que esta idea estratégica que acabamos de ilustrar es la base de nuestro sentido de la responsabilidad personal y moral.)

Existen, no obstante, personas que no responden a la ley como se desearía. Hay un requerimiento tácito de que la ley debe ser inteligible; esto es, directa en su expresión y fácil de entender. Es lógico suponer entonces que las personas que se encuentran bajo su jurisdicción la comprendan, pero para algunas tal suposición implica una exigencia desmesurada. Me refiero a los casos paradigmáticos de los insanos y los mentalmente incapaces. Los eximimos de toda responsabilidad legal porque el poder disuasorio de la ley no los alcanza ni siquiera mínimamente y el intento de educarlos, elevándolos hasta el umbral del entendimiento y del conocimiento, sería estéril (o al menos muy costoso). Castigarlos como si fueran ciudadanos responsables socavaría la institución misma del castigo

[7] Una vez trazada la línea, es dable esperar que los jueces actúen con discriminación, pues cabe que se produzcan desequilibrios: un juez *podría* considerar que la ignorancia del acusado es debida a circunstancias atenuantes —"'ignorancia justificada de la ley" (Gross 1979, pág. 155)— pero, en general, el juez no está obligado a considerar el punto.

(que depende de su credibilidad) al socavar su fundamento racional. Negarse a admitir que estas personas no son responsables es ofender la ley y la racionalidad de la ciudadanía; como también se ofende a la justicia expidiendo y ejecutando leyes que no son del conocimiento público. De ese modo, agregamos disposiciones explícitas que eximen a ciertas personas de toda responsabilidad legal con el objeto de preservar la credibilidad y la defensa del sistema.

Como resultado de ello, disminuye la cantidad de posibles condenados, cuya selección se realizará entre los auténticamente responsables y los legalmente culpables. Pero quizá los principios usados para delimitar esta categoría sean demasiado groseros y se necesiten distinciones más sutiles si lo que se pretende es una mayor aceptación y credibilidad de la ley; es decir, si lo que se pretende es perfeccionar la justicia de la ley. En esta instancia cabe, sin embargo, la propuesta subversiva: ¿por qué no exculpar a *todos* los que cometen delitos? Si la ley no pudo disuadirlos en esa ocasión, entonces eran incapaces de ser disuadidos —al menos en esa ocasión.

Pero dar un paso semejante comportaría el aniquilamiento de nuestro sistema "disuasorio". Se disiparía el efecto de contención que leyes y sanciones ejercen sobre las deliberaciones de los ciudadanos (eventualmente) responsables y volverían los indeseables males inherentes al estado de naturaleza. Por lo tanto, debemos ser arbitrarios una vez más y trazar la línea —cuya importancia, en cuanto a la ubicación exacta, es la misma que en el caso de establecer la edad legal para consumir bebidas alcohólicas o conducir automóviles. Necesitamos un umbral de competencia legal que pueda determinarse con eficacia, no suponiendo que no puede haber "contraejemplos" intuitivamente persuasivos a la línea que tracemos, estipulando previamente que no se tendrán en cuenta esas cosas. Es necesario no demorarse en los ínfimos detalles de las circunstancias del acusado, sino establecer rápida y concretamente si el agente es capaz de ser disuadido *en general*, aunque no lo fuera en el momento del delito. (Este es el ámbito de la libertad de acción, en la justicia y en la moral, que buscábamos en el capítulo 6 para refutar el actualismo.)[8]

[8] Fingarette (1972, pág. 203) define la insania como "el fracaso para responder con relevancia a lo que es esencialmente relevante [...] en virtud de una deficiencia grave de la *capacidad* para responder de ese modo, inherente a la índole mental de la persona, al menos durante ese lapso". Según Honoré (véase capítulo 6) se debe comprender a la persona como a alguien que ya no "puede" (en el sentido general del verbo) responder con relevancia. Si por el contrario, interpretamos el verbo poder en sentido particular, toda demostración de locura ante testigos (por ejemplo, agredir al oficial de policía que lo arresta) constituiría una evidencia *definitiva* de insania (¿transitoria?). La idea de insania transitoria provoca, sin duda, muchos recelos y objeciones. Ello se debe al oscuro reconocimiento de que se trata de un delgado límite que podría dar paso a exculpaciones indiscriminadas y, por lo tanto, subversivas.

Nuestra negativa a ahondar las causas y circunstancias más allá de un cierto límite arbitrario puede infundir sospechas. ¿Se trata de una política justa? En realidad, sí. Recuérdese que el azar se promedia. Sería posible mejorar el baloncesto descalificando los aciertos fortuitos y las casualidades desafortunadas. (En mi opinión, no hay mejor juego en la ciudad que el de *afirmar que la gente es responsable*.) Recuérdese, que en el fundamento racional de la insuficiencia de la represión legal está implícita una cierta tolerancia al delincuente que se arriesga. Si los malhechores, luego de considerar las posibilidades, deciden que el delito bien vale la pena el riesgo, lo cometen y son atrapados, ello no es un signo *prima facie* de irracionalidad. Del hecho de que se perdió una apuesta no podemos concluir que fue hecha irracionalmente. Si la gente se arriesgó con pleno conocimiento de las consecuencias del fracaso, no puede alegar ahora que la sanción aplicada es injusta.

Una institución con un fundamento racional de esa naturaleza está destinada a crear —o a *constituir*— una clase de agentes legalmente culpables cuya ulterior sujeción al castigo mantiene la credibilidad de las sanciones. Si la institución ha de perdurar, tendrá que ajustar convenientemente sus umbrales arbitrarios en la medida en que aparezca nueva información (o información errónea) que socave su crédito. Cabe pensar que existe una posición óptima del umbral de competencia que maximiza el efecto de contención de las leyes (para todo tipo de circunstancias sociales, grados de complejidad pública, etc.). Un umbral más alto que el óptimo fomentará en los acusados una suerte de simulación que, al ser identificada por el pueblo, disminuirá el respeto por las leyes y su poder de disuasión. Un umbral más bajo que el óptimo producirá una disminución del poder disuasorio de la ley y llevará a castigar a individuos que, ante los ojos de la sociedad, "realmente no podían evitarlo". La opinión pública acerca de la justicia de la ley constituye un factor crítico para la eficacia del derecho.

El esquema que acabo de esbozar es una versión de la defensa ordinaria de la ley, el juicio y el castigo en tanto que institución disuasoria (existen otras versiones que he preferido omitir). A la luz de las circunstancias familiares, presenta el mismo tipo de fundamento racional que el de una "institución" mucho más esencial, si bien históricamente posterior: "el mundo de los conceptos morales" de la responsabilidad individual.

¿Por qué nos consideramos a nosotros mismos y a los demás como *moralmente* responsables? A primera vista, esta pregunta puede parecer parasitaria del problema metafísico acerca de las condiciones que se necesitan para *ser* verdaderamente responsable. (Cuando se critica una cierta posición de los umbrales de respon-

sabilidad *legal*, se puede aducir que la ley considera responsables a individuos que no son moralmente responsables de su comportamiento y viceversa.) Es posible establecer una distinción entre la pregunta acerca de por qué *consideramos* que las personas son responsables o *asumimos* la responsabilidad de varias cosas y la pregunta acerca de si somos realmente responsables y, en ese caso, cuál es la razón. No ignoramos a veces que se cometen errores: se considera responsable a quien no lo es —ya sea por el hecho en cuestión o por sus acciones en general—; una persona que es responsable se rehúsa a asumir su responsabilidad.

Pero sea lo que fuere la responsabilidad, considerada como un estado metafísico, no tendrá derecho a reclamar nuestra atención en tanto no sea posible relacionarla con algún *desideratum* social reconocible. ¿Qué puede importarnos si nos hallábamos o no en ese estado metafísico cuando cometimos un acto reprobable? Desde luego, los seres humanos son capaces de desearlo todo y el anhelo por ese tipo de responsabilidad bien puede surgir cuando se está dispuesto a satisfacer una apetencia puramente metafísica. (Imagínese que alguien se las ingeniara para sentir el deseo de comer un trozo de pan compuesto por las mismas moléculas que componían el trozo de pan que alguna vez comió Alejandro Magno. Ahora, imagínese a un ser humano que se las ingenia para desear fervientemente la responsabilidad metafísica.)

¿Por qué, pues, necesitamos afirmar que las personas —incluso nosotros mismos— son responsables? "Al sostener que una persona es responsable y obrar en consecuencia podemos conducirla a que se desprenda de un rasgo indeseable; y esto es útil, independientemente de que el rasgo sea o no obra suya" (Gomberg 1978, pág. 203). Una vez más se evidencia que una cuota de arbitrariedad es siempre útil. En lugar de indagar, indefinidamente, con el propósito de *descubrir* si un determinado rasgo de carácter es obra de la persona, en lugar de verificar exactamente hasta qué punto un determinado yo se construye a sí mismo, simplemente *consideramos* a las personas responsables de su conducta (dentro de ciertos límites que procuramos no examinar muy detenidamente). La recompensa que obtenemos al adoptar esta estrategia es una mayor proporción de comportamiento "responsable" que de ese modo se inculca.

Veamos las implicaciones de esta perspectiva respecto de la pregunta que a menudo nos formulamos acerca de nuestra culpabilidad o inocencia.¿Acaso nos engañamos al *asumir* la responsabilidad de nuestros delitos? Si se postulara la existencia de una especie de responsabilidad cósmica, absoluta, en la que uno participa, o bien de la cual está excluido, entonces la perspectiva de rechazar o asumir erróneamente responsabilidades se transformaría en un

problema complejo y en apariencia insoluble. Peor aun, adoptando la posición kantiana, nos veríamos obligados a admitir que la única conducta responsable sería la conducta moralmente perfecta; en ese caso, la culpabilidad sería tan imposible como la cuadratura del círculo. Wolf nos ofrece una formulación actual de esa concepción.

> Afirmo con los compatibilistas que la responsabilidad moral de un agente no depende de si está determinado o no sino de cómo está determinado... Sin embargo, puesto que pienso que la única determinación satisfactoria es la que se ejerce en virtud de las razones que un agente debe poseer, se desprende de ello que un agente puede estar determinado y ser responsable al mismo tiempo, pero sólo en tanto y en cuanto ejecute las acciones que debe ejecutar. Si, por otra parte, ejecuta una acción moralmente mala, entonces la acción no puede estar determinada correctamente. (Wolf 1980, pág. 163.)

Según esta teoría, debe existir algo malo y criticable en el diseño o en la implementación de la secuencia causal de razonamientos que conduce a una acción moralmente mala.

Pero hemos visto que esto no es necesariamente así. En cada nivel de organización, desde el nivel de organización de la memoria hasta el del diseño de las intituciones sociales, los *mejores diseños posibles,* dadas las restricciones impuestas por la finitud y la presión del tiempo, incluyen una cierta cuota de arbitrariedad y de riesgo prudente. La organización (enteramente inconsciente) de la memoria garantiza que sólo nos sucederá un subconjunto de puntos relevantes en el tiempo disponible (Cherniak 1983). Todo estilo de autocontrol desarrollado por un individuo obtiene su eficacia a expensas del riesgo de decidir arbitrariamente entre las diversas opciones que se le presentan. A veces, la solución de un problema o la toma de decisiones debe incluir una decisión arbitraria (consciente o no) para terminar con las deliberaciones y llegar a la decisión final, aunque se tenga pleno conocimiento de que todavía quedan por examinar importantes consideraciones. Y en el plano de las instituciones sociales encontramos no sólo la arbitrariedad del derecho penal (en situaciones estratégicas), sino también otros casos de arbitrariedad, como por ejemplo el uso de sorteos para acelerar, correcta y oportunamente, decisiones, que sería imposible tratar de otra manera (decisiones como, por ejemplo, quiénes serán reclutados en las fuerzas armadas). Todo sistema finito de control (como el cerebro humano) es propenso a cometer errores o a tomar decisiones que condenaría un análisis más cuidadoso. Se trata de un rasgo inevitable del carácter humano, incluso del más perfecto: el pecado original naturalizado. Sin embargo, es prudente adoptar políticas que minimicen las consecuencias negativas de estos inevitables

defectos de carácter, empleando fuerzas correctivas similares a las que bosquejamos en nuestro fundamento racional del castigo en el derecho. Si se afirma (con cierta arbitrariedad) que las personas son responsables de sus actos y se toman recaudos para que comprendan que se las considerará como tales, se reducirá a límites tolerables el riesgo del delito en el diseño (y el rediseño) de sus caracteres. Cuando, a pesar de estas medidas, se atrapa a un individuo en el acto mismo de delinquir, simplemente habrá perdido sus apuestas (razonables o no) y no deberá objetar la pena que se le asigne.

En efecto, ésta es la actitud que se toma comúnmente cuando uno enfrenta las propias transgresiones (al menos las más leves). Hace poco *olvidé* sacar el permiso de estacionamiento en una zona restringida y asumí la responsabilidad de pagar una buena multa. Ciertamente, no pude atenuar la pena alegando fragilidad de memoria o "una distracción transitoria más allá de mi control". Aun cuando alguna de estas explicaciones fuera verdadera, sería difícil encontrar un caso más evidente de culpabilidad por pecado (venial) de omisión.

Cuando los riesgos son mayores y los delitos más graves, ya no es tan común la inmediata aceptación de la culpabilidad. No obstante, son muchos los delincuentes que no dudan de sus culpas y, si bien esta admisión es, en ocasiones, un autoengaño alimentado por la sociedad, no siempre es necesariamente así. El escepticismo respecto de la posibilidad misma de la culpabilidad proviene de la veneración a un ideal absoluto: el concepto de culpa ante los ojos de Dios. Pero el hecho de que esa condición no se cumpla en el mundo no debe inducirnos al escepticismo respecto de la integridad de la institución de la responsabilidad moral.

3. La negación del secreto terrible

Es difícil ser optimista en las frías mañanas, cuando reflexionamos sobre nuestra abrumadora debilidad y estulticia, conscientes del abismo que separa nuestra imagen pública de nuestros pensamientos privados vergonzantes que en apariencia determinaron un acto. ¿Cuál es la reacción correcta o al menos coherente ante la confusión de esos lamentables defectos de carácter? ¿El autorreproche?

> ¿Cuál es la manera habitual de hacernos reproches? Decirnos, por ejemplo: "¡qué negligente!"; "¡qué irreflexivo!"; "¡qué egoísta!"; "¡qué torpe y desenfrenado!". "¿Cómo he podido, siquiera por un instante, considerar las cosas con tanta pequeñez y exasperación?" En otras palabras, nos atribuimos, respecto del hecho y sólo en cierta medida,

> un carácter malo y lo lamentamos. En esto consiste todo autorreproche. Dirigimos contra nosotros mismos la desaprobación y el disgusto; deseamos deshacer lo hecho en el pasado y ante la imposibilidad, nos embarga una rabia y una vergüenza exacerbadas *por tener eso adentro*, eso que nos condujo a perpetrar lo que ahora lamentamos. (Hobart 1934, pág. 4.)

¿Pero la contrición, los remordimientos y la autocondena son acaso menos incoherentes que el deseo frustrado de deshacer lo hecho? Una vez que hemos visto la utilidad social del mito del libre albedrío, podemos preguntarnos si estamos obligados por alguna otra razón a instituirlo en el dominio privado y a considerarnos a nosotros mismos responsables. ¿Cuál es la actitud alternativa? ¿La apesadumbrada contrición, la impasibilidad del espectador o la rabia ante la injusticia de estar determinado de antemano a ser un miembro despreciable de la raza humana?

¿Cómo deberíamos reaccionar ante la propia culpabilidad? Si el concepto de culpa que se aplica es el tradicional, absoluto"ante los ojos de Dios", entonces será tan descalificable como las dudosas nociones que están emparentadas con él: la perfecta voluntad kantiana, el yo sartreano que se crea a sí mismo, el agente ideal socrático que nunca hace el mal a sabiendas, el agente de Chisholm como motor inmóvil. Nadie, ni siquiera monstruos como Hitler o Eichmann, ni criminales comunes y cuerdos como Agnew o Vesco, ni usted cuando quebrantó la ley o una promesa, pueden ser culpables en ese sentido; la tradición filosófica y teológica ha restringido tanto ese sentido de culpa, que es imposible describir el estado que el término pretende denominar.

Desde este punto de vista, el hecho de no ser culpable no es un gran consuelo. Pero cabe todavía la responsabilidad legal, sea en sentido literal o en sentido relativo que comporta el afirmar que uno es personal y moralmente responsable. ¿Hay espacio para el remordimiento y la contrición en estas nociones? La pregunta es sumamente compleja y merece cuidadoso análisis (véase Strawson 1962 y Dennett 1980), pero aventuraré algunas consideraciones bastante obvias. La conciencia que se manifiesta a sí misma sólo retrospectivamente, en una agonía de remordimientos, pero que jamás lo hace prospectivamente, tratando de superar algún impulso primario, constituye un rasgo de carácter nada atractivo y completamente prescindible en cualquier mundo moral. Pero el remordimiento que es manifestación de un cambio significativo y no transitorio de las prioridades que atañen a la decisión es precisamente el tipo de actitud que sólo puede ponerse en práctica gracias a la afirmación de nuestra responsabilidad. Por lo tanto, hay espa-

cio para el remordimiento en nuestro concepto naturalizado de culpa que establecimos.

Pero supóngase que, al considerar la propia biografía, uno cae en la cuenta de que —al margen de sus buenas cualidades— fue, es y será un villano cuya vida habrá de transcurrir en medio de dificultades (difamado, condenado, castigado y evitado por sus congéneres). Hay personas así. Como agentes, han aprobado el examen de responsabilidad pero son tan desalmados e inaccesibles a toda calidez humana que merecen nuestro desprecio. ¿Somos injustos? La réplica es la siguientes: ¿quién merece ser másd despreciado sino aquel que es completamente despreciable? ¿Acaso pueen evitar el desprecio de sus semejantes? No *pueden* evitarlo *ahora*, pues han ido demasiado lejos (tal como el borracho responsable de haber hecho de sí mismo un borracho); sin embargo, en el capítulo 4 hemos visto que uno es tan responsable de su carácter como de todo lo que es fruto de sus esfuerzos pasados. Si la persona se ha convertido en un monstruo, entonces no quedará nadie a quien responsabilizar, per ése es el caso límite.

¿Cómo consolar o aconsejar a quien se confiesa despreciable? La respuesta es obvia: si se está empeñado en"salvar su alma", se alentará a la persona, de la manera más efectiva posible, a modificar sus actitudes y a luchar por superar su mala reputación (bien merecida). Si lo que se desea es paliar sus tribulaciones (sin tener en cuenta el costo social), se la inducirá a percibir su absoluta degradación e impotencia con el propósito de fomentar una actitud de apatía y fatalismo, logrando así (quizá) que su reflexión se disocie casi por completo de sus actos y proyectos y que adopte una actitud de cínica tolerancia hacia su lado peor.

Todos hemos experimentado alguna vez el terrible daño existencial en el cual reconocemos que hemos caído en la actitud paralizante del espectador pasivo que se pregunta inútilmente lo que pensará o hará en el futuro inmediato. En lugar de considerar el futuro con esperanzas, de hacer planes o tratar de anticiparnos al mundo, nos precipitamos en una espiral regresiva que nos aniquila al sumirnos en una honda preocupación por nosotros mismos que malgasta nuestro tiempo y que prácticamente nos garantiza que se hará realidad la imagen de futilidad e indecisión que nos obsesiona. Afortunadamente para la mayoría, las depresiones pasan pronto y al cabo volvemos a contraer compromisos con la sociedad. Y emergemos del marasmo como el golfista que finalmente ha captado la sabiduría del consejo: mantener la cabeza baja hasta el final del movimiento.

Puesto que todos conocemos ese sentimiento, podemos comprender lo ominoso de toda demostración o descubrimiento que

pretenda probar la imposibilidad del libre albedrío. Tener buenas razones para desear el libre albedrío no es obviamente lo mismo que tener buenas razones para creer que se tiene libre albedrío. Sin embargo, parece que tener buenas razones para desear el libre albedrío es tener buenas razones para tratar de creer que se tiene libre albedrío. En efecto, como hemos visto, es muy probable que el hecho de creer que se tiene libre albedrío sea una de las condiciones necesarias para tener libre albedrío: un agente que gozara de las otras condiciones necesarias —racionalidad y capacidad de autocontrol y de introspección de orden superior—, pero que fuera inducido engañosamente a creer que carece de libre albedrío, estaría tan inhabilitado por dicha creencia para elegir libre y responsablemente como por la falta de cualquiera de las otras condiciones.

Esto coloca en una posición delicada al agnóstico que todavía no está seguro de que se cumplan las demás condiciones necesarias (¿las hemos identificado a todas?) Si puede superar o ignorar sus dudas y lograr el estado de ánimo a un creyente sincero del libre albedrío, logrará uno de los siguientes estados: a) un estado genuino de libre albedrío, o b) al menos la ilusión de libre albedrío.

El primer estado es, evidentemente, el más deseable pero, a falta de otra cosa, también el último *podría serlo*, ya que pertenece a la clase de ilusiones que hacen posible o estimulan la vida: la ilusión de que aún nos ama la persona amada; la ilusión de que nos quedan muchos años de vida, aunque no sea así; la ilusión de que la fealdad física no impide que los otros perciban nuestra belleza interior.

Si, por el contrario, el agnóstico se sume en la duda o, peor aún, se convence de la ausencia de libre albedrío, su actitud hacia las oportunidades de elección o de acción será tal que lo inhabilitará esencialmente como elector y dejará de obrar según la idea de libertad. Pero, por naturaleza, somos aparentemente incapaces de mantener esa convicción por más de unos pocos malos momentos. El hambre, el aburrimiento o la curiosidad nos sobrevienen al cabo y nos conducen otra vez a la acción (una similitud con la *sphex* que deberíamos agradecer). Sin embargo, la experiencia, aunque breve, es fascinante y sus implicaciones lo son aún más si se las toma seriamente en cuenta. No nos sorprenden que sobren motivos para mirar el lado luminoso del asunto y concluir que el libre albedrío constituye un problema insoslayable y apremiante. Pero las circunstancias favorecen el autoengaño y toda defensa del libre albedrío contra el escepticismo será sospechosa, en el mejor de los casos, de mera expresión de deseos y en el peor de hipocresía.

No obstante, aquello que deseamos que sea verdadero puede serlo; y dadas las aspiraciones de racionalidad, la irrefrenable curiosidad y el escepticismo que padecemos, no nos queda otro camino que un paciente examen del problema con vistas a determinar si se sostiene la anhelada conclusión. Esto es lo que hice y la conclusión es optimista: el libre albedrío no es una ilusión, ni siquiera una de esas ilusiones irreprimibles que nos ayudan a vivir. Cuando observamos con cuidado el origen de nuestros miedos y recelos, constatamos una y otra vez que no se trata de axiomas irrefutables o de descubrimientos empíricos bien fundamentados, sino de imágenes fuera de foco, apenas vislumbradas, como las sombras que en la pared del cuarto adquieren una consistencia amenazadora justamente porque no las observamos de cerca.

Cuando deseamos el libre albedrío, deseamos, en realidad, el poder de decidir nuestros actos y de decidirlos con sabiduría, a la luz de expectativas y deseos. No queremos que nadie nos controle sino ser agentes capaces de tomar la iniciativa y la responsabilidad de nuestros actos y proyectos. He procurado demostrar que todo ello nos pertenece porque es el producto natural de nuestros dones biológicos, ampliados y reforzados por nuestra iniciación en la sociedad.

Por otra parte, deseamos suficiente libertad de acción para ejercer esos poderes en el mundo, sin necesidad de acomodarlos permanentemente al único, desesperado, curso de acción que tiene la posibilidad de satisfacer nuestros deseos. Esta libertad también puede ser accesible y bien vale luchar por ella, pero no está garantizada. Existen reales limitaciones a la libertad humana que no son de índole metafísica: por ejemplo, las obligaciones políticas, la represión, la manipulación ejercida a través de la difusión de información falsa y la desesperante "jugada forzosa" de la pobreza y del hambre. Podríamos, sin duda, combatir tales restricciones si no estuviéramos condenados a usar otra clase de chaleco de fuerza: el curioso límite creado (e impuesto) por el ejercicio mismo de la libertad y por el hecho mismo de reconocernos responsables; es decir, las cadenas, sogas y cuerdas y todos los arreos concomitantes del compromiso (tácito o explícito) que contraemos con la familia, los amigos, los proyectos de vida y que nos vuelven cada vez más incapaces de responder al llamado de la libertad radical.

Nuestras experiencias con estas amenazas reales y variables a la libertad generan angustia y dan origen a las metáforas que motivan y guían (en secreto) los argumentos y los análisis de los filósofos. Mi estrategia ha consistido en examinar, en primer término, esas metáforas y descubrir el papel que les toca en los experimentos mentales —los generadores de intuición— que dominaron nuestra imaginación durante largo tiempo.

Nos asusta la idea de que la ciencia descubra, esté por descubrir o amenace con descubrir que nunca seremos lo que queremos ser. La amenaza no proviene del determinismo —si fuera así, podríamos suspirar aliviados pues los físicos están, aparentemente, de acuerdo sobre la naturaleza indeterminista del mundo— sino de la ciencia misma y de su visión de las cosas: el"naturalismo".

Este miedo ha sobrevivido, como la mayoría de los miedos, por ignorancia. Lo alimenta una idea en exceso simplista de cuanto la ciencia tiene que decir sobre la causalidad, el tiempo, la posibilidad y sobre nosotros mismos y el resto del universo. En tanto nos neguemos a examinar detenidamente cuál podría ser la imagen científica de la humanidad —por miedo a lo que podríamos encontrar—, subsistirá la sospecha de que los argumentos filosóficos abstractos que pretenden demostrar la compatibilidad de la libertad y de la ciencia no pasan de ser meros tanteos en la oscuridad. Después de todo, las amenazas asumen formas muy diferentes.

He procurado exorcizar a un ogro por vez —aunque algunos vuelan bajo apariencias distintas—, demostrando que los argumentos tan persuasivos que los invocan se originan, virtualmente enn todos los casos, en el uso indebido de los generadores de intuición, cuando la atención se centra en modelos simples cuya misma simplicidad es la que genera las intuiciones ilusorias.

La táctica de juzgar qué es posible e imposible, basándose en lo que uno puede o no concebir o imaginar, constituye una parte ineliminable del método filosófico y del método científico. Se dice: "es inconcebible para mí que *p*" y luego se abrevia: "es inconcebible que *p*" o, en otras palabras: "como todos pueden observar, *p* es imposible". ¿Usted dice que no puede imaginar que *p* y en consecuencia declara que *p* es imposible? ¿No se trataría de una *hybris*? Una de mis tácticas para responder las tradicionales declaraciones filosóficas sobre lo inimaginable, ha sido conminar a poner más empeño.

Mis propios generadores de intuición se diseñaron para facilitar la tarea. En efecto, si ponemos más empeño, podemos imaginar un ser que escucha la voz de la razón y, sin embargo, no está exento de causalidad. En efecto, podemos imaginar un ser cuyas decisiones son causadas por la interacción de los elementos de su estado actual con los elementos del entorno sobre los cuales no tiene ningún control y, sin embargo, *tiene el control* y no está controlado por ese entorno omnipresente y omnicausal. En efecto, es posible imaginar un proceso de autocreación que comienza con un agente no responsable y crea gradualmente un agente responsable de su propio carácter. En efecto, es posible imaginar un ser racional y *determinista* que no se engaña cuando considera que el futuro es "abierto" y "depende de

él". En efecto, es posible imaginar un agente responsable y libre con repecto al cual sea verdadero que toda vez que actuó en el pasado, no pudo hacerlo de otra manera.

No sólo es posible imaginar las cosas que se suponían inimaginables, socavando así los argumentos que afirman su imposibilidad, sino que la misma perspectiva de imaginarlas en detalle produce dividendos. Se perfilan con más nitidez las ideas de racionalidad, autocontrol, autoría, oportunidad, prevención y evitación, responsabilidad y autoperfeccionamiento que se creían amenazadas. Cuando vemos cuán sólida puede ser la visión naturalizada de nosotros mismos, somos menos propensos a dejarnos seducir por las teorías absolutistas y perfeccionistas que se niegan a asimilar la ciencia o el sentido común. Una versión más realista de la racionalidad humana —que tome en cuenta el problema de la finitud— será más versátil y poderosa que las teorías exageradamente idealizadas que dieron origen a todas las paradojas tradicionales.

Sé que la posición naturalista a la que adhiero, la que nos alienta a imaginarnos como robots orgánicos, o como partes diseñadas del universo, resulta odiosa para muchos humanistas. He procurado demostrarles que al apartarse de la visión naturalista le dan la espalda a una fuente inagotable de ideas *filosóficas*. Este tema se halla implícito en las citas y en las notas al pie de página referidas a trabajos cuyas áreas de competencia suelen ser ignoradas por los filósofos (especialmente por los que escriben sobre el libre albedrío). Pienso que una consideración empática de los frutos de esas disciplinas brinda una perspectiva fácilmente accesible para quien sólo haya estudiado la literatura filosófica tradicional. Y he comprobado que mi incompetencia para sumergirme profundamente en esas disciplinas técnicas, extrañas y a menudo prohibidas, ha frustrado mi curiosidad filosófica en innumerables ocasiones. Los filósofos con la capacidad y el entrenamiento apropiados pueden sacar gran provecho del estudio de tales disciplinas.

Habría, ciertamente, mucho más que decir sobre el libre albedrío. He revisado todo el material pertinente hasta nuestros días y, desde luego, no he tomado en serio algunas ideas que otros pueden considerar fundamentales. Hay quienes se sentirán consternados ante mi rotunda descalificación (en el capítulo 6) del "problema" de reconciliar el libre albedrío con la hipótesis de ser omnisciente, infinito. Esta versión tradicional del libre albedrío constituyó, durante siglos, el centro del problema, pero su interés depende de una concepción muy literal y antropomórfica de Dios como sapiente. No creo que en la actualidad se pueda considerar a Dios de esa manera, por ello decidí que no vale la pena discutirla. En mi opinión, sólo se trata de un enigma histórico o de un entretenimiento intelectual y

no he querido dar la impresión (en parte, porque me he divertido mucho al escribir este libro) de que tomo las cosas a la ligera.

Desde luego, cabe esperar objeciones a nuestra convicción de que tenemos libre albedrío, y serán bienvenidas pues lo que nos interesa, en definitiva, es conocer la verdad. Los incompatibilistas, "los deterministas duros" y otros escépticos inventarán argumentos nuevos e ingeniosos para demostrar que, en realidad, nadie tiene, o podría tener, libre albedrío si el determinismo, el mecanicismo y otros "ismos" son verdaderos. Terminaré este libro con algunos consejos para refutar estas objeciones.

En primer lugar, examine detenidamente qué clase de libre albedrío pone en peligro el argumento. ¿Vale la pena preocuparse por ella? Pregúntese si tiene razones bien fundamentadas para creer que posee esa clase de libre albedrío, o bien para temer que no la posee. Si le faltara ese tipo de libertad, ¿se sentiría como un prisionero o como una marioneta? Quizá la conclusión del nuevo argumento sólo se refiera a la imposibilidad de poseer una propiedad metafísica que, en el mejor de los casos, sólo les interesa a los académicos. O más aun: el anhelo de la libertad en cuestión puede resultar a la luz de la investigación una bendición incoherentemente concebida. Pregúntese si puede imaginar seres cuya voluntad sea más libre que la nuestra. En ese caso, ¿de qué rasgos lamentables, que forman parte de nuestra dote como organismos físicos, estarían exentos? Si el ideal de libertad que le quita el sueño es contradictorio en sus términos, no debería apesadumbrarlo el hecho de que sea inalcanzable. No tiene sentido estrujarse las manos porque no se puede deshacer el pasado, ni evitar lo que sucede en el presente, ni crearse a sí mismo *ex-nihilo,* ni elegir ambas alternativas en una sola decisión, ni ser prefecto. Si el argumento en cuestión pasa todas estas pruebas y pareciera avizorarse en lontananza alguna conclusión temible, entonces y sólo entonces se lo tomará en cuenta y se averiguará exactamente de qué argumento se trata.

Bibliografía

Anscombe, G. E. M. (1957): *Intention*, Oxford, Blackwell.

Austin, J. L. (1961): "Ifs and Cans", en Austin: *Philosophical Papers*, comp. por J. O. Urmson y G. Warnock, Oxford, Clarendon Press, págs. 153-180.

Ayer, A. J.(1954): "Freedom and Necessity", en *Philosophical Essays*, Nueva York, Macmillan, págs. 271-284. [Hay versión castellana: *Ensayos filosóficos*, Barcelona, Ariel, 1979.]

Ayers, M. (1968): *The Refutation of Determinism: an Essay in Philosophical Logic*, Londres, Methuen.

Bennett, J. (1976): *Linguistic Behaviour*, Cambridge, Cambridge University Press.

Bennett, J. (1980):"Accountability", en van Straaten 1980.

Berlin, I. (1954): *Historical Inevitability*, Oxford, Oxford University Press.

Berofsky, B., comp. (1966): *Free Will and Determinism*, Nueva York, Harper & Row.

Berry, M. (1983):"Breaking the Paradigms of Classical Physics From Within", en J. Petitot, comp.: *Proceedings of the Cerisy Symposium (Logos et Theorie des Catastrophes).*

Boër, S. y W. Lycan (1980): "Who, Me", *The Philosophical Review*, 89, págs. 427-466.

Brewer, W. F. (1974): "There is No Convinging Evidence for Operant or Classical Conditioning in Adult Humans", en W. B. Weimer, comp.: *Cognition and the Symbolic Processes*, Hillsdale, Nueva Jersey, Erlbaum.

Brown, N. (1959): *Life Against Death*, Nueva York, Random House.

Campbell, D. (1975): "On the Conflicts Between Biological and Social Evolution and Between Psychology and Moral Tradition", *American Psychologist*, 30, págs. 1103-1126.

Cartwright, H. (1970): "Quantities", *The Philosophical Review*, 79, págs. 25-42.

Chaitin, G. (1976): "Randomness and Mathematical Proof", *Scientific American*, 232, págs. 47-52.

Chappell, V. (1962): *The Philosophy of Mind*, Englewood Cliffs, NJ, Prentice-Hall.

Cherniak, C. (1983): "Rationality and the Structure of Human Memory", *Synthese*, 57, págs. 163-186.

Chisholm, R. (1961): "Responsability and Avoidability", en Hook 1961.

Chisholm, R. (1964a): "J. L. Austin's Philosophical Papers", *Mind*, 73, págs. 1-26.

Chisholm, R. (1964b): "Human Freedom and the Self", The Lindley Lecture, 1964, University of Kansas. (Reimpreso en Watson 1982.)

Chisholm, R. (1976): *Person and Object: A Metaphysical Study*, The Paul Carus Lectures. Lasalle, Ill., Open Court.

Cohen, L. J. (1981): "Can Human Irrationality be Experimentally Demonstrated?", *Behavioral and Brain Sciences*, 4, págs. 317-331.

Cummins, R. (1980): "Culpability and Mental Disorder", *Canadian Journal of Philosophy*, X, págs. 207-232.

Danto, A. (1965): "Basic Actions", *American Philosophical Quarterly*, 2, págs. 141-148.

Davidson, D. (1970): "How is Weakness of the Will Possible?", en Joel Feinberg, comp., *Moral Concepts*, Oxford, Oxford University Press, págs. 93-113.

Davidson, D. (1973): "On the Very Idea of a Conceptual Scheme", *Proceedings and Addresses of the American Philosophical Association*, 1973-1974, 67, págs. 5-20.

Dawkins, R. (1976): *The Selfish Gene*, Oxford, Oxford University Press. [Hay versión castellana: *El gen egoísta*, Barcelona, Labor, 1979.]

Dawkins, R. (1980): "Good Strategy or Evolutionary Stable Strategy", en G. W. Barlow y J. Silverberg, comps.: *Sociobiology: Beyond Nature/ Nurture?*, AAAS Selected Symposium, Boulder, Colorado, Westview Press.

Dawkins, R. (1982): *The Extended Phenotype*, Nueva York, W.H. Freeman.

Dennett, D. (1969): *Content and Consciousness*, Londres, Routledge & Kegan Paul, y Nueva York, Humanities Press.

Dennett, D. (1972): Reseña de J. R. Lucas: *The Freedom of the Will*, en J. Phil, 67, págs. 91-97.

Dennett, D. (1973): "Mechanism and Responsability", en Hinderich 1973. (Reimpreso en Dennett 1978a.)

Dennett, D. (1975): "Why the Law of Effect Will not Go Away", *Journal of the Theory of Social Behaviour*, 5, págs. 169-187. (Reimpreso en Dennett 1978a.)

Dennett, D. (1976): "Conditions of Personhood", en A. Rorty 1976. (Reimpreso en Dennett 1978a.)

Dennett, D. (1978a): *Brainstorms: Philosophical Essays on Mind and Psychology*, Montgomery, Vt., Bradford Books, y Hassocks, Sussex, Harvester.

Dennett, D. (1978b): "Current Issues in the Philosophy of Mind", *American Philosophical Quarterly*, 15, págs. 249-261.

Dennett, D. (1980): "The Milk of Human Intentionality" (comentario acerca de Searle 1980), *Behavioral and Brain Sciences*, 3, págs. 428-430.

Dennett, D. (1981a): "Three Kinds of Intentional Psychology", en R. Healey, comp.: *Reduction, Time and Reality*, Cambridge, Cambridge University Press, págs. 37-61.

Dennett, D. (1981b): "True Believers: The Intentional Strategy and Why It Works", en A. F. Heath, comp.: *Scientific Explanation*, Oxford, Oxford University Press.

Dennett, D. (1982a): "Beyond Belief", en A. Woodfield, comp.: *Thought and Object*, Oxford, Oxford University Press.

Dennett, D. (1982b): "Notes on Prosthetic Imagination", *New Boston Review*, junio, págs. 3-7. (Reimpreso, con revisiones, como "The Imagination Extenders", en *Psychology Today*, diciembre de 1982, págs. 32-39.)

Dennett, D. (1982c): "Why do we think what we do about why we think what we do?" (comentario acerca de Goodman 1982), *Cognition*, 12, págs. 219-237.

Dennett, D. (1982d): "How to Study Consciousness Empirically: or Nothing Comes to Mind", *Synthese*, 53, págs. 159-180.

Dennett, D. (1982e): "Comments on Rorty", *Synthese*, 53, págs. 349-356.

Dennett, D. (1983a): "Styles of Mental Representation", *Proc. Aristotelian Soc.*, Nueva Serie, 83, 1982-1983, págs. 213-226.

Dennett, D. (1983b): "Intentional Systems in Cognitive Ethology; the 'Panglosian Paradigm' Defended", en *Behavioral and Brain Sciences*, vol. 6, págs. 343-390.

Dennett, D. (en preparación-a): "Cognitive Wheels: The Frame Problem of Artificial Intelligence", en C. Hookway, comp.: *Minds, Machines and Evolution*, Cambridge, Cambridge University Press.

Dennett, D. (en preparación-b): "The Self as a Center of Narrative Gravity", en P. M. Cole, D. L. Johnson y F. S. Kessel, comps.: *Self and Consciousness*, Nueva York, Praeger.

Descartes, R.: *Oeuvres de Descartes*, edición de C. Adam y P. Tannery, París, Leopold Cerf, 1897-1913. [Hay múltiples versiones castellanas de las obras más importantes de Descartes]

De Sousa, R. (1970): "Self-Deception", *Inquiry*, 13, págs. 308-321.

Dewey, J. (1922): *Human Nature and Conduct*, Nueva York, Henry Holt. [Hay versión castellana: *Naturaleza humana y conducta*, México, Fondo de Cultura Económica.]

Doan, B. y J. Macnamara: *Free Will and Scientific Insight*, en preparación.

Dretske, F. (1981): *Knowledge and the Flow of Information*, Cambridge, Mass., Bradford/MIT Press.

Dyson, F. (1979): *Disturbing the Universe*, Nueva York, Harper. [Hay versión castellana: *Trastornando el universo*, México, Fondo de Cultura Económica.]

Eccles, J. C. (1953): *The Neurophysiological Basis of Mind*, Oxford, Oxford University Press.

Edwards, P. (1961): "Hard and Soft Determinism", en Hook 1961.

Eells, E. (1982): *Rational Decision and Causality*, Cambridge, University Press.

Everett, H. (1973): En *The Many-Worlds Interpretation of Quantum Mechanics*, comp. por B. S Dewitt y N. Graham, Princeton, N. J., Princeton University Press.

Farrell, B. A. (1950): "Experience", *Mind*, 1950. (Reimpreso en Chappell 1962.)

Fingarette, H. (1972): *The Meaning of Criminal Insanity*, Berkeley, University of California Press.

Fischer, J. (1982): "Responsability and Control", *J. Phil.*, enero de 1982, págs. 24-40.

Fishman, G.S. (1973): *Concepts and Methods in Discrete Event Simulation*, Nueva York, Wiley. [Hay versión castellana: *Conceptos y métodos en la simulación digital de eventos discretos*, México, Limusa, 1978.]

Flew, A. (1955): "Divine Omniscience and Human Freedom", en A. Flew y A. MacIntyre, comps.: *New Essays in Philosophical Theology*, Londres, SCM Press.

Fodor, J. (1980): "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", *Behavioral and Brain Sciences*, 1980, vol. 3. (Reimpreso en Haugeland 1981 y en J. Fodor 1981.) Representations, Cambridge, Mass., Bradford/MIT Press.

Fodor, J. (1983): *The Modularity of Mind*, Cambridge, Mass., Bradford/MIT Press. [Hay versión castellana: *La modularidad de la mente*, Madrid, Morata, 1986.]

Foot, P. (1973): "Nietzsche: the Revaluation of Values", en R. Solomon, comp., *Nietzsche: a Collection of Critical Essays*, Nueva York, Doubleday, págs. 156-168.

Ford, J. (1983): "How Random is a Coin Toss?", *Physics Today*, abril de 1983, págs. 1-8.

Frankfurt, H. (1969): "Alternate Possibilities and Moral Responsibility", *J. Phil.*, 65, págs. 829-833.

Frankfurt, H. (1971): "Freedom of the Will and the Concept of a Person", *J. Phil.*, 68, págs. 5-20. (Reimpreso en Watson 1982.)

Frankfurt, H. (1973), "Coercion and Moral Responsibility", en Honderich 1973.

Frankfurt, H. (1976):"Identification and Externality", en A. Rorty 1976.

Gale, G. (1981): "The Anthropic Principle", *Scientific American*, 254, diciembre, págs. 154-171.

Gallup, G. (1977): "Self-recognition in Primates: A Comparative Approach to Bidirection Properties of Consciousness", *American Psychologist*, 32, págs. 329-338.

Gallwey, Y. (1979): *The Inner Game of Tennis*, Nueva York, Bantam.

Gazzaniga, M. y J. Ledoux (1978): *The Integrated Mind*, Nueva York, Plenum.

Gibbs, J. P. (1975): *Crime, Punishment and Deterrence*, Nueva York/Oxford/ Amsterdam, Elsevier.

Gibson, J. J. (1966): *The Senses Considered as Perceptual Systems*, Boston, Houghton Mifflin.

Gibson, J. J. (1979): *The Ecological Approach to Visual Perception*, Boston, Houghton Mifflin.

Gomberg, P. (1978): "Free Will as Ultimate Responsability", *American Philosophical Quarterly*, 15, pág. 208.

Goodman, N. (1965): *Fact, Fiction and Forecast*, Nueva York, Bobbs-Merrill.

Goodman, N. (1982): "Thoughts Without Words", *Cognition*, 12, págs. 205-218.

Greenspan, P. (1976): "Wiggins on Historical Inevitability and Incompatibilism", *Philosophical Studies*, 29, págs. 235-247.

Greenspan, P. (1978): "Behavior Control and Freedom of Action", *Phil. Review*, 87, págs. 25-40.

Gross, H. (1979): *A Theory of Criminal Justice*, Nueva York, Oxford University Press.

Hacking, I. (1975): "All Kinds of Possibility", *The Philosophical Review*, 84, págs. 321-337.

Hampshire, S. (1959): *Thought and Action*, Londres, Chatto and Windus.

Hart, H. L. A. (1968): *Punishment and Responsability*, Oxford, Oxford University Press.

Haugeland, J. (1981): "Semantic Engines: an Introduction to Mind Design", en Haugeland, comp.: *Mind Design*, Cambridge, Mass., Bradford/MIT Press.

Hintikka, K. J. J. (1973): *Logic, Language-Games and Information*, Oxford, Clarendon. [Hay versión castellana: *Lógica, juegos de lenguaje e información*, Madrid, Tecnos, 1976.]

Hintikka, K. J. J. (1976): "The Semantics of Questions and the Questions of Semantics", *Acta Philosophica Fennica*, vol. 28, #4, Amsterdam, North-Holland.

Hobart, R. E. (1934): "Free Will as Involving Determinism and as Inconceivable Without It", *Mind*, 43, págs. 1-27.

Hobbes, T. (1841): *The Questions Concerning Liberty, Necessity, and Chance*, en *The English Works of Thomas Hobbes*, vol. V, edición de sir W. Molesworth, Londres, John Bohn. (Extractado en Morgenbesser and Walsh 1962.)

Hofstadter, D. R. (1979): *Gödel, Escher, Bach: an Eternal Golden Braid*, Nueva York, Basic Books.

Hofstadter, D. R. (1981): "Reflections" (sobre Searle), en Hofstadter and Dennett 1981, págs. 373-382.

Hofstadter D. R. (1982a): "Undercut, Flaunt, Hruska, Behaivoral Evolution and Other Games of Strategy", *Scientific American*, 247, agosto de 1982, págs. 16-24.

Hofstadter, D. R. (1982b): "Can Creativity be Mechanized?", *Scientific American*, 247, septiembre 1982, págs. 18-34.

Hofstadter, D. R. (1982c): "Variations on a Theme as the Essence of Imagination", *Scientific American*, 247, octubre de 1982, págs. 20-29.

Hofstadter, D. R. (1983): "The Calculus of Cooperation is Tested Through a Lottery", *Scientific American*, 248, junio 1983, págs. 14-28. (Véase también M. Gardner: "The Topology of Knots and the Results of Douglas Hofstadter's Luring Lottery", *Scientific American*, 249, septiembre de 1983, págs. 25-28.)

Hofstadter, D. R. y Dennett, D. (1981): *The Mind's I: Fantasies and Reflections on Mind and Soul*, Nueva York, Basic Books.

Honderich, T., comp. (1973): *Essays on Freedom and Action*, Londres, Routledge and Kegan Paul.

Honoré, A. M. (1964): "Can and Can't", *Mind*, 73, págs. 463-479.

Holmstrom, N. (1977): "Firming up Soft Determinism", *The Personalist*, 58, págs. 39-51.

Hook, S., comp. (1961a): *Determinism and Freedom in the Age of Modern Science*, Nueva York, Collier.

Hook, S. (1961b): "Necessity, Indeterminism and Sentimentalism", en Hook 1961a.

Hospers, J. (1961): "What Means This Freedom?", en Hook 1961a.

Hume, D. (1739): *A Treatise of Human Nature*, edición de L. A. Selby-Bigge, Oxford, Clarendon, 1964. [Hay versión castellana: *Tratado de la naturaleza humana*, Barcelona, Orbis, 1984.]

Humphrey, N. (1982): "Consciousness: a Just So Story", *New Scientist*, 95, pág. 475.

Hutchings, E. (1983): "The Autonomous Viking", *Science*, 219, 18 de febrero de 1983, págs. 803-808.

Jacob, F. (1982): *The Possible and the Actual*, Seattle, University of Washington Press.

James, W. (1921): "The Dilemma of Determinism", en W. James, *The Will to Believe*, Nueva York, Longmans, Green, págs. 145-183.

Jauch, J. M. (1973): *Are Quanta Real?*, Bloomignton, Indiana, Indiana University Press.

Jaynes, J. (1976): *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, Boston, Houghton Mifflin.

Jennings, P. (1962): *The Jenguin Pennings*, Harmandsworth, Penguin.

Kahneman, D., P. Slovic y A. Tversky, comps. (1982): *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.

Kant, I. (1788): *Critique of Pure Reason* (trad. de K. Abbot), Londres, Longmans, 1873. [Hay diversas versiones castellanas, entre otras: *Crítica de la razón pura*, Madrid, Alfaguara, 1984, 3ª ed.]

Krebs, J. R., A. Kacelnik y P. Taylor (1978):"Test of Optimal Sampling by Foraging Great Tits", *Nature*, 275, págs. 27-31.

Kyburg, H. (1983): "Rational Belief", *Behavioral and Brian Sciences*, 6, págs. 231-274.

Laplace, P. (1820): *A Philosophical Essay on Probabilities*, trad. de F. W. Truscott y F. L. Emory, Nueva York, Dover, 1951.

Lem, S. (1974): "The Seventh Sally or How Trurl's Own Perfection Led to No Good", en *The Cyberiad*, trad. de M. Kandel, Nueva York, Seabury Press. (Reimpreso en Hoftsdter y Dennett 1981.)

Levi, I. (1980): *The Enterprise of Knowledge*, Cambridge, Mass., MIT Press.

Locke, J. (1690): *An Essay Concerning Human Understanding* (edición de A. C. Fraser), Nueva York, Dover, 1959. [Hay versión castellana: *Ensayo sobre el entendimiento humano,* Madrid, Aguilar, 1961, 2ª ed.]

Lucas, J. R. (1961): "Minds, Machines and Gödel", *Philosophy*, 36, págs. 112-127. (Reimpreso en A. R. Anderson, comp.: *Minds and Machines*, Englewood Cliffs, N.J., Prentice-Hall, 1964.)

Lucas, J. R. (1970): *The Freedom of the Will*, Oxford, Clarendon.

Lucrecio Caro, J., *The Nature of the Universe*, trad. de R. E. Latham, Hardmondsworth, Penguin, 1951. [Hay versión castellana: *De la naturaleza* (Obra completa), Madrid, Consejo Superior de Investigaciones, 1983, 2ª ed.]

MacIntyre, A. (1957): "Determinism", *Mind*, 66, págs,. 28-41.

MacKay, D. M. (1960): "On the Logical Indeterminacy of a Free Choice", *Mind*, 69, págs. 31-40.

Mackie, J. L. (1974): *The Cement of the Universe*, Oxford, Oxford University Press.

McDermott, D. (1982): "A Temporal Logic for Reasoning about Processes and Plans", *Cognitive Science*, 6, págs. 101-155.

Milgram, S. (1974): *Obedience to Authority: An Experimental view*, Nueva York, Harper and Row. [Hay versión castellana: *Obediencia a la autoridad; un punto de vista experimental,* Bilbao, Desclée de Brouwer.]

Mill, J. S. (1867): *An Examination of Sir William Hamilton's Philosophy*, Londres, Longmans, Green, Reader and Dyer. (El capítulo XXVI, "On the Freedom of the Will", fue reimpreso en Morgenbesser and Walsh 1962.)

Monod, J. (1972): *Chance and Necessity*, Nueva York, Knopf.

Morgenbesser S. y J. Walsh (1962): *Free Will*, Englewood Cilffs, N. J., Prentice-Hall.

Nagel, T. (1974): "What is it Like to be a Bat?", *Philosophical Review*, 83, págs. 435-450. (Reimpreso en Nagel 1979 y, con un comentario, en Hofstadter y Dennett 1981.)

Nagel, T. (1979): "Moral Luck", en Nagel, *Mortal Questions*, Cambridge, Cambridge University Press.

Nagel, T. (1981): "Freedom and Objectivity", manuscrito no publicado fechado el 6 de julio de 1981, citado con autorización.

Neisser, U. (1976): *Cognition and Reality*, San Francisco, Freeman.

Nietzsche, F. (1883): *Thus Spoke Zarathustra*, trad. de W. Kaufmann, en *The Portable Nietzsche*, edición de W. Kaufmann, Nueva York, Viking. [Hay diversas versiones castellanas, entre otras: *Así hablaba Zaratustra*, Madrid, Edaf, 1982, 12ª ed.]

Nietzsche, F. (1887): *On the Genealogy of Morals*, trad. de W. Kaufmann y R. J. Hollingdale, Nueva York, Vintage, 1967.

Nozick, R. (1979): "Coercion", en S. Morgenbesser, P. Suppes y M. White, comps.: *Philosophy, Science and Method: Essays in Honor of Ernest Nagel*, Nueva York, St. Martin's Press, págs. 440-472.

Nozick, R. (1981): *Philosophical Explanation*, Cambridge, Mass., Cambridge University Press.

Pears, D. (1964): "Predicting and Deciding", *Proc. of the British Academy*, 50, 1964, págs. 193-227. (Reimpreso en P. F. Strawson, comp.: *Studies in the Philosophy of Thought and Action*, Oxford University Press, 1968, págs. 97-134).

Popper, K. (1951): "Indeterminism in Quantum Physics and Classical Physics", *British Journal for the Philosophy of Science*, I, págs. 179-188.

Popper, K. y J. Eccles (1977): *The Self and Its Brain*, Nueva York, Springer. [Hay versión castellana: *El yo y su cerebro*, Barcelona, Labor, 1982.]

Powers, L. (1978): "Knowledge by Deduction", *Phil. Review*, 87, págs. 337-371.

Putnam, H. (1975): "The Meaning of 'Meaning'", en K. Gunderson, comp.: *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Sciences, Vol. VII, Minneapolis, Univ. Of Minnesota Press. (Reimpreso

en Putnam: *Mind, Language and Reality*, Cambridge, Cambridge University Press, 1975.)

Pynchon, T. (1973): *Gravity's Rainbow*, New York, Viking.

Quine, V. W. O. (1960): *Word and Object*, Cambridge, Mass., MIT Press.

Raab, F. V. (1955): "Free Will and the Ambiguity of Could" , *Phil. Review*, 64, págs. 60-67.

Rabin, M. (1980): "Probabilistic Algoritms in Finite Fields", *SIAM Journal of Computing*, 9, págs. 273-280.

Rawls, J. (1971): *A Theory of Justice*, Cambridge, Mass., Harvard University Press. [Hay versión castellana: *Teoría de la justicia*, México, Fondo de Cultura Económica, 1979.]

Rawls, J. (1980): "Representation of Freedom and Equality", Dewey Lectures, III, *J. Phil.*, vol. 77, págs. 535-559.

Reid, T. (1788): *Essays on the Active Power of the Human Mind*, Cambridge, Mass., edición del MIT Press, 1969.

Robbins, T, (1976): *Even the Cawgirls Get the Blues*, Nueva York, Bantham Books.

Rorty, A. (1972): "Belief and Self-Deception", *Inquiry*, 15, págs. 387-410.

Rorty, A. (1976): *The Identities of Persons*, Berkeley, University of California Press.

Rorty, R. (1979): *Philosophy and the Mirror of Nature*, Princeton, Princeton University Press. [Hay versión castellana: *Filosofía y el espejo de la naturaleza*, Madrid, Cátedra.]

Rorty, R. (1982): "Contemporary Philosophy of Mind", *Synthese*, 53, págs. 323-348.

Ryle, G. (1949): *The Concept of Mind*, Londres, Hutchinson. [Hay versión castellana: *El concepto de lo mental*, Buenos Aires, Paidós.]

Ryle, G. (1954): *Dilemmas*, Cambridge, Cambridge University Press. [Hay versión castellana: *Dilemas*, Inst. de Investigaciones Filosóficas, México, UNAM, 1979.]

Ryle, G. (1970): "Autobiographical", en O. P. Wood y G. Pitcher, comps.: *Ryle: a Collection of Critical Essays*, Garden City, Nueva York, Doubleday Anchor.

Ryle, G. (1979): *On Thinking*, edición de K. Kolenda, Oxford, Blackwell.

Saarinen, E., comp. (1979): *Game-Theoretical Semantics*, Dordrecht, Reidel.

Sanford, D. (1975): "Infinity and Vagueness", en *Phil. Review*, 84, págs. 520-535.

Sartre, J. P. (1946): *L'Existentialisme est un Humanisme: Dialogue avec P. Naville*, París, Nagel. [Hay versión castellana: *El existencialismo es un humanismo*, Barcelona, Orbis, 1985, 2ª ed.]

Sartre, J. P. (1956): *Being and Nothingness*, traducido al inglés por H. Barnes, Nueva York, Philosophical Library. [Hay versión castellana: *El ser y la nada*, Madrid, Alianza, 1984.]

Schiffer, S. (1976): "A Paradox of Desire", *American Philosophical Quarterly*, 13, págs. 195-212.

Schneewind, J. (1981): Reseña de John Stuart Mill en *Autobiography and Literary Essays*, comp. por J. M. Robson y J. Stillinger (*Collected Works of John Stuart Mill*, vol. I) en MLN, vol. 96, dic. de 1981, págs. 1231-1235.

Schotch, P. K. y R. E. Jennings (1980): "Modal Logic and the Theory of Modal Aggregation", *Philosophia*, 9, págs. 265-278.

Seidenfeld, T. (1981): "Levi on the Dogma of Randomization in Experiments", en R. J. Bogdan, comp., *Henry E. Kyburg, Jr. and Isaac Levi*, Dordrecht, Reidel, págs. 263-291.

Searle, J. (1980): "Minds, Brains and Programs", *Behavioral and Brain Sciences*, 3, págs. 417-157.

Sellars, W. (1963): *Science, Perception and Reality*, Londres, Routledge & Kegan Paul.

Sher, G. (1979): "Effort, Ability and Personal Desert", *Philosophy and Public Affairs*, 8, págs. 361-376.

Simon, H. (1957): *Models of Man*, Nueva York, Wiley.

Simon, H. (1969): *The Sciences of the Artificial*, Cambridge, Mass., MIT Press.

Skinner, B. F. (1953): *Science and Human Behavior*, Nueva York, Macmillan. [Hay versión castellana: *Ciencia y conducta humana*, Barcelona, Martínez Roca.]

Skinner, B. F. (1971): *Beyond Freedom and Dignity*, Nueva York, Knopf. [Hay versión castellana: *Más allá de la libertad y la dignidad*, Barcelona, Martínez Roca.]

Slote, M. (1982): "Selective Necessity and Free Will", *J. Phil.*, 74, págs. 5-24.

Sober, E. (1981): "The Evolution of Rationality", *Synthese*, 46, págs. 95-120.

Spinoza, B. (1677): *Ethics*, trad. inglesa de R. H. M. Elwes, Londres, George Bell & Sons, 1891. [Hay versión castellana: *Etica*, México, Fondo de Cultura Económica, 1958.]

Straight, S. (1977): "Consciousness as a Workspace", *SISTM, Quarterly*, 1, págs. 11-14.

Strawson, P. F. (1959): *Individuals*, Londres, Methuen.

Strawson, P. F. (1962): "Freedom and Resentment", *Proc. of the British Academy*, 1962. (Reimpreso en Strawson 1968.)

Strawson, P. F., comp. (1968): *Studies in Philosophy of Thought and Action*, Oxford, Oxford University Press.

Taylor, C. (1976): "Responsibility for Self", en A. Rorty 1976. (Reimpreso en Watson 1982.)

Taylor, R. (1964): "Deliberation and Foreknowledge", *American Philosophical Quarterly*, 1, págs. 73-80. (Reimpreso en Berofsky 1966.)

Thorp, J. (1980): *Free Will: a Defence Against Neurophysiological Determinism*, Londres y Boston, Routledge & Kegan Paul.

Ullman, S. (1979): *The Interpretation of Visual Motion*, Cambridge, Mass., MIT Press.

Van Inwagen, P. (1975): "The Incompatibility of Free Will and Determinism", *Phil. Studies*, 27, págs. 185-199. (Reimpreso en Watson 1982.)

Van Inwagen, P. (1978): "Ability and Responsibility", *Philosophical Review*, 87, págs. 201-224.

Van Inwagen, P. (1983): *An Essay on Free Will*, Oxford, Clarendon.

Van Straaten, Z., comp. (1980): *Philosophical Subjects: Essays Presented to P. F. Strawson*, Oxford University Press.

Von Neumann, J. y O. Morgenstern (1947): *Theory of Games and Economic Behavior*, 2ª edición, Princeton, Princeton University Press.

Von Wright, G. H. (1974): *Causality and Determinism*, Nueva York, Columbia University Press.

Watson, G., comp. (1982): *Free Will*, Oxford, Oxford University Press.

Whitaker, H. (1976): "A Case of the Isolation of Language Function", en H. Whitaker y H. A. Whitaker, comps., *Studies in Neurolinguistics*, vol. II, Nueva York, Academic Press.

Wiener, N. (1948): *Cybernetics: or Control and Communication in the Animal and the Machine*, Cambridge, Mass., Technology Press y Nueva York, John Wiley & Sons. [Hay versión castellana: *Cibernética*, Barcelona, Tusquets.]

Wiggins, D. (1973): "Towards a Reasonable Libertarianism", en Honderich 1973.

Williams, B. (1973): *Problems of the Self*, Cambridge, Cambridge University Press.

Williams, B. (1981): *Moral Luck*, Cambridge, Cambridge University Press.

Wimsatt, W. (1980): "Randomness and Perceived-Randomness in Evolutionary Biology", *Synthese*, 43, págs. 287-329.

Wittgenstein, L. (1953): *Philosophical Investigations*, trad. de G. E. Anscombe, Oxford, Blackwell. [Hay versión castellana: *Investigaciones filosóficas*, Grijalbo, Barcelona.]

Wittgenstein, L. (1967): *Zettel*, edición de G. E. M. Anscombe y G. H. von Wright, Oxford, Basil Blackwell. [Hay versión castellana: *Papeles*, México, UNAM, 1979.]

Wolf, S. (1980): "Asymmetrical Freedom", *J. Phil.*, LXVII, págs. 151-165.

Wolf, S. (1981): "The Importance of Free Will", *Mind*, 190, págs. 368-378.

Wolff, R. P. (1973): *The Autonomy of Reason: a Commentary on Kant's Groundwork of the Metaphysic of Morals*, Nueva York, Harper & Row.

Wooldridge, D. (1963): *The Machinery of the Brain*, Nueva York, McGraw Hill.

Wooldridge, D. (1968): Mechanical Man: *The Physical Basis of Intelligent Life*, Nueva York, McGraw Hill.

Zatchick, A. (1977): "On Deserving to Deserve", *Philosophy and Public Affairs*, 6, págs. 370-388.

Indice temático

Términos diposicionales, 144
Distracción, 188
Agente absoluto, 102, 149
Absolutismo, 44n., 50, 59, 99, 101, 102, 148-149, 171, 172, 179, 180, 186-189, 194
Abstracta, 38, 59
Accidente, 118, 122, 164, 170, 174
Reconocimiento, 107
Acción, 13, 93, 130, 166, 187
básica, 130n.
causada, 47
vs movimiento, 21
Acontecimientos, independencia de, 162, 170-171, 173-174
Actual *vs* posible, 121
Actualismo, 165, 166, 171, 184
Adaptacionismo, 74n., 166n.
Adicto, 125
"Agenciedad", 27, 34, 69, 92, 95, 98n., 167, 168, 175
verbos de, 144, 145
Agente, 19-25, 32, 44n., 49, 51, 59, 63, 72, 78n., 80-81, 88-96, 101-104, 106-107n., 122, 129n., 133, 134, 137, 141, 142, 146, 148, 151, 154, 157, 163, 177-180, 185, 187, 190, 192
causalidad del, 93, 98, 175
secreto, 53
o cuasiagente, 71
control sin, 73, 77, 79
Aglomerativo, 170
AGNEW, S., 189
Avión (de juguete), 69, 71, 76, 95n. *Véase también* Piloto.
Akrasia, 125. *Véase también* Voluntad, falta de.
Alquimia, 43
Modalidad alética, 169
ALEJANDRO MAGNO, 186
Algoritmo, 40-41, 84, 87, 171-172
factible, 172
aleatorio, 172-173
"consideradas todas las cosas", 52, 87, 125
Posibilidades alternativas, principio de las, 113
Altruismo, 57, 61
Amplificación,
de efectos, 94
de la suerte, 111
de efectos cuánticos, 139, 156; *véase también* Contador Geiger.
Analogía, 17-21, 27, 29-30, 32, 109
Ananké, 165
Angeles, 180, 181
Anima, 26
ANSCOMBE, G. E. M., 28, 46n., 197
Hormiga, 128
"hormiga" como término de masa, 128
Oso hormiguero, 128
Antropomorfismo, 36, 78, 194
Anticuerpos, 147
Antiintuitivo, 35, 172; *véase también* Intuición.
Ansiedad, 17-20, 32, 59, 120, 192
Anzio, desembarco de, 39
Apatheia, 14
Apatía, 14, 27, 122, 124, 190
Afasia, 56

A priori, razonamiento, 64, 101
Arbitrario, 14, 85-88, 183-188
ARISTÓTELES, 42, 95n., 125
Armas, carrera cognitiva de, 53
Arte, 15-16, 87, 109
  de la autodefinición, 98-110
Inteligencia artificial, 44, 54, 55, 87n., 132, 147
Atomos, 14, 173-174
  desvío de los, en el cerebro, 14
  de oxígeno e hidrógeno, 166, 173
Actitud,
  en las relaciones interpersonales, 28n.
  participativa, 28n., 63
  reactiva, 28n.
  del espectador, 190
Agustiniano, 165
AUSTIN, J. L., 33, 47n., 68n., 151, 167-169, 197
Autoría, 108
Autómatas, 23, 48n.
  teoría de los, 68
Autoestimulación, 56-58
Promedio, 128
  la suerte se promedia, 113-116, 140, 174, 185
Evitar, 17, 36, 70, 71, 79, 116-117, 122, 142-150, 174, 194
AYER, A. J., 48, 49, 63n., 82n., 163n., 197
AYERS, M., 141, 143, 166, 169 n., 197

Malo, 36
BAIER, K., 105n.
Baloncesto, 113-116, 147
Murciélago, 31
Abejas, danza de las, 54
Cerveza, jarro de, 33
Comportamiento, 40-41, 45-46, 49
  comprensión meramente conductal, 50-54, 57
Conductismo, 73-74
Creencia, 42, 49, 54, 74, 78, 81-82, 104
  en el determinismo, 26-27, 120-123
Benevolencia de la naturaleza, 73-74
BENNETT, J., 27, 58, 63n., 189, 197
Tiempo bergsoniano, 129n.
Berkeley, máquina de, 137n.
BERLIN, I., 20n., 148n., 197
BEROFSKY, B., 197
BERRY, M., 173, 197
Billar, bolas de, 48, 77
Biología, 24, 37, 40, 51, 141, 162, 166-171, 192

  "biológico" como sinónimo de "genético", 47n.
  trascender la, 60-61
BIRD, A., 113n.
Pájaros, 26, 40-41, 83n., 147, 174
  insectívoros, 128, 174
  cerebro de los, 40-41
Culpa, 27, 49, 118, 153, 180-181
Tabula rasa, 100
Ciego, 51
  y arbitrario, 14
  causación ciega, 46, 48
  extracción ciega del número de la lotería, 139
  saltos, 104, 107
  ceguera ante nuestros propios procedimientos, 149
  punto, 126
  ensayo y error, 44-45
"Inglés corporal", 29
Temperatura corporal, mantenimiento de la, 36
BOËR, S., 99n., 197
Ogros, 19, 20, 22, 23, 33, 77, 82, 140, 151
  del determinismo genético, 47n.
Obligaciones, políticas, 192
Soluciones "de libro", 103, 105, 106-107n.
BORGES, J. L., 92n.
Boston, 71, 91
BRADFORD, J., 118
Cerebro, 14, 42-44, 46, 80-81, 156
  daño cerebral, 26
  electrodos en el, 20-21, 152
  monitor del, 137
  tamaño del cerebro de la *sphex,* 37n.

tumor cerebral, 20-21
BRAMHALL, Dr., 34
BREWER, W., 26
BROWN, N. O., 86, 197
Causación bruta, 24, 149
Fuerza bruta, 136
algoritmo de la, 172
Buridán, asno de, 86n.

CAMPBELL, D., 61n., 100n., 197
"Poder", 33, 68-69, 151, 165-174
particular y general, 168, 184n.
Pena capital, 13
Caricatura, 19
Preocupación,
por el control, 68n., 132
por saber si "podría haber obrado de otra manera", 151-164
por el libre albedrío, 16-19
CARTWRIGHT, H., 128n., 198
Estado catatónico, 123n.
Error categorial, 92
Conocimiento, teoría causal del, 74n.
Causalidad,42, 46-49, 94, 122, 124, 127, 193
ausencia o exención de la, 59, 84, 157n.
bruta o cruda, 24, 46-49
y control, 69-73, 75-76, 77, 145, 160
trama de la, 92-93, 120, 148, 159
ideas de la causalidad en la filosofía griega, 13-14
como necesidad, 82n., 143, 148
Causación, *véase* Causalidad.
División celular, 141
Universo, cemento del, 159
Cuartel central, 96-97
Certeza cartesiana, 149
Ceteris paribus, 146
CHAITIN, G., 84, 138, 172, 198
Azar, 77, 85, 135, 138-140, 165, 177
Cambio, 128, 144, 147, 159, 163
grados de, 133
Caótico, 128-129, 130, 148, 172n.
sistemas caóticos en la física newtoniana, 173
CHAPPELL, V., 31n., 198
Carácter, 49, 101, 105, 109, 113, 158-162, 165, 193
CHERNIAK, C., 87n., 187, 198
Ajedrez, 87-90, 107
Niño,
cósmico, 21-22
niños, 101, 103, 118
Cuarto chino, 31
CHISHOLM, R., 93, 102, 143, 167, 189,198
Elección, 14, 34n., 45, 61, 124; *véase también* Decisión;
radical, 107-108
Cristianismo, 176
CIA, 38
Ciudadanos, 114
Psicología clínica, 47
Armario sin fantasma, 23
Bruto, 173
Coerción, 78, 81, 192
"Cogito ergo sum", 31
COHEN, L. J., 26n., 198
Arrojar la moneda, 26n., 77, 84, 85, 109-112, 140, 167
Coincidencia, 76, 162, 163n., 171; *véase también* Accidente;
absoluta, 171
Comatoso, 37, 48
Cometa, 145
Comunicación, 54-57, 60
Compatibilismo, 14, 127, 151-152, 187, 193
Competencia, 19, 23, 57, 77-82
Complejidad, 25
Comprensión, 39-40, 42-43, 50, 56-57
Compulsión, 78, 101
Computabilidad, teoría de la, 46
Ordenador, 52, 72, 88, 137n., 138-139, 172
ciencia de la computación, 51
Apuestas, juegos de, 111n.
El concepto de lo mental, 19
Esquema conceptual, 63, 130, 134
Condicionamiento, 49
operativo, 26
Conflicto, entre yo y yo mismo, 108-109
Conciencia, 50-61, 96, 153, 175-176, 186, 187, 189

acceso a las decisiones de la, 97
excentricidad de la, 96
de la necesidad, 82n.
autoconciencia, 52-53
Consideración, 45, 58, 89, 103-107, 128, 138, 164, 187
Proceso constitutivo, 36, 59-62, 64, 108-109, 185
Contracausal, 61, 163n.
Control, 20-24, 32, 96, 99-100, 101-105, 112, 119, 125, 130, 135, 137, 145, 153, 161, 188, 192-193
experimentos de control en el laboratorio, 48, 70, 76
remoto, 68-72
tener el control, 72n.
autocontrol, 32, 66-90, 92, 98-101, 104, 112-113, 117, 119, 131,138, 141, 146, 165, 169, 187, 191-192, 194
skinneriano, 75, 83n.
sistemas de, 36, 43, 61
teoría del, 83n., 130-131
en experimentos mentales, 81
transitividad del, 72
Cooperación, 106-107n.
Coordinación, problemas de, 54
Cuerpo calloso, 56n.
Corteza, 92
Costos y beneficios, 37n., 86, 128, 162, 187
Contrafáctico, 124n., 129n.
Creatividad, 109-110
Acreedor, 176
CRICK, F., 38n.
Crimen, 180-184, 188-189
Criminal interior, 92
CUMMINS, R., 180, 198
Curador, no guardián, 37

DANTO, A., 130n.
DARWIN, Ch., 38, 44, 59, 78, 124
DAVIDSON, D., 63, 125, 198
DAWKINS, R., 24n., 35n., 37n., 47n., 55n., 61, 72n., 82n., 141, 198
Decisión, 27, 45, 47, 80-82, 86-88, 97, 100-102, 119, 124-125, 130-133, 135-136, 190, 192-193
como causa de la acción, 42
causada, 25
tomar decisiones *vs* dejar que sucedan, 13
moral, 116
fenomenología de la, 95n.
procedimiento de, 107, 172; *véase también* Algoritmo.
teoría de la, 97n.
árbol de, 87
Ciervo, 20
Artículo, definido, 147
Libertad, grados de, 69, 130, 162
Deliberación, 23, 28, 71, 85-89, 104, 117-118, 120-136, 139, 143-144, 146-150, 152, 159-160, 164, 169, 184, 187
DEMÓCRITO, 140
Demonio,
maligno de Descartes, 31, 70n., 148
de Laplace, 129, 139, 173
DENNETT, D., 25, 26n., 31n., 37n., 38n., 43, 45n., 46n., 50n., 54n., 57n., 70n., 72n., 78, 87n., 92n., 96, 99n., 105n., 110n., 129n., 137n., 144n., 147, 163n., 166n., 199-200n.
DESCARTES, R., 9n., 27, 31, 34, 69, 87, 149
Diseño, 26, 37, 40, 43, 45, 48-49, 59, 64, 78, 86, 88, 100, 127, 129, 132-134, 138, 147, 160-163, 170, 174, 187
Deseo, 20, 53, 68, 71, 73, 77, 81, 101, 104, 192
del controlador, 68, 71, 73-76, 77, 130-131, 137
de orden superior, 105n.
"implantado", 49
suspendido, 51
DE SOUSA, 125, 200
Destino, 13, 27, 112, 140, 171
Determinismo, *passim*,
definición del, 13
genético, 47n.
duro, 101, 195
blando, 101
Disuasión, 181-185

Diablo, 23
DEWEY, J., 17, 200
Dados, 14, 77, 77n.
Razón, dictados de la, 34, 35, 39-42, 45, 57, 106, 107n., 149, 152
Establecer una diferencia, 29, 35, 121-122, 135, 144-145, 147-148, 159
Dignidad, 175-178
Responsabilidad disminuida, 33, 177,178
Tretas sucias, 81, 89, 139
Yo que desaparece, 26, 91-98
Discriminación, 41, 42-43
Desorden, 83-90, 149
Perturbar el universo, 33, 144, 147, 148
DOAN, B., 47, 200
Perro,
  puede ladrar, 129-130
  atado a un carro, 14, 20
Hacer *vs* suceder, 93, 96-98
Muñecos, 21
Dominó, 93, 119
"No importa", 162
DOSTOYEVSKI, F., 27
Dibujo *vs* escultura, 15
Secreto terrible, 27-28, 188-195
DRETSKE, F., 76n., 86, 200
Dualismo, 42, 148
Deber, 175
Programación dinámica, 40, 41
DYSON, F., 33, 200
Tierra, 20, 71, 72, 85, 89, 145, 147, 160, 181
ECCLES, J., 19n., 200
Eclipse, 173
EDWARDS, P., 101, 102, 109, 116, 200
EEG (electroencefalograma), 137
EELLS, E., 97n., 200
EICHMANN, A., 189
EINSTEIN, A., 25, 137
Elba, 20
Libertad de acción,
  entre lo real y lo posible, 165, 166, 184
  para el control, 79, 103
  para la deliberación, 33, 132
  como margen para el error, 89, 102-103
  para la ley y la moralidad, 184
  para el yo, 26-27
Electrodos, en el cerebro, 20
Ingeniería, 51, 71, 160n., 162-164, 166
Engaño, 38, 53
  coloración engañosa, 38
  del genio maligno de Descartes, 70n.
  autoengaño, 27, 33, 126, 180, 187, 191
Entropía, 35
Entorno, 35, 44, 65, 77, 100, 158
  control ejercido por el, 73-76, 78, 100
  características del, 128, 137
Epicúreos, 13-14
Epistémico, 85n., 148, 168-170, 173-174
  intereses, 82
  posibilidad, 129, 132, 168-169, 173
  problema del control, 69, 79-81
  problema de la autoidentificación, 99
  tratamiento, 76
  giro, 146n.
Equilibrio, actos en, 81
Libre albedrío, característica esencial del, 68n.
Esencialismo, 50n.
Recurrencia eterna, 22
Esperanza, en los replicantes, 14
Evaluación radical, 108, 119
EVERETT, H., 144n., 159, 200
Evolución, 36-39, 45, 53-61, 65, 75, 78, 88-89, 100, 139, 166-167, 170-171, 177n.
  estrategias evolutivamente estables, 82n.
  como certamen de selección natural, 112
Exculpación, 118, 153, 179-181, 185, 188
Excusas, 164
Existencialismo, 107
  conflicto existencial, 139
  temor existencial, 190
Expectativas, 130-131, 174, 192
  y control, 70-71

Experimentos, 76; *véase también* Experimento mental.
Explicación, 146
Representación explícita de las razones, 38-40, 45
Exploración,
vs explotación, 40
pautas de, 138, 149, 173
en la percepción, 54
Externalización, 164
Aruspicia, 86n.
Manchas oculares, 38
Facultad, la razón como una, 39n.
Incapacidad, 138, 161, 168, 184n.
Justo. *Véase* Injusto.
FARRELL, B. A., 31, 201
Fatalismo, 17, 29, 143, 149-150, 190
local, 123-125, 136, 148n., 149, 150
Miedo, 17-21, 24, 27-29, 68, 75, 80, 82n., 88, 177, 179, 192-193
de volar, 153
traficantes de, 17, 20, 24, 29
Retroalimentación, circuitos de, 67n., 79-80, 89, 188
Cuidarse a sí mismo, 36
Fermat-Lagrange, teorema de, 172
FINGARETTE, H., 184, 201
Finitud, 88, 106, 127, 129n., 134, 138, 178, 187
el problema de la, 194
Perspectiva, de la primera persona, 95
FISCHER, J., 21, 67n., 79n., 80, 152, 201
FISHMAN, G. S., 201
Adaptación, 129
FLEW, A., 68n., 201
FODOR, J., 43n., 63, 64, 201
Follow-through, 29-30, 126, 190
Búsqueda del alimento, 36
FOOT, P., 62, 201
Carreras, 113-144
Jugadas forzadas, 88, 90, 192
FORD, J., 84, 173, 201
Previsión (y conocimiento previo) 70-71, 78, 79, 82, 88, 89, 105, 127,143
Ver el bosque, 25
Perdón, 180
Formalización prematura, 15
FORSTER, E. M., 96
Fundamentalismo, 130
Problema de encuadre, 147
FRANKFURT, H., 73, 104n., 105n., 152, 153, 201
Libertad, *passim*, ,17-18, 34n., 35, 52
política no es metafísica, 17
estoica como resignación, 14
Fundamentos lógicos flotantes, 38, 82n., 83n., 88, 176
Rana en el fondo del jarro de cerveza, 33, 151, 167
Futilidad, 123, 126, 190
Futuro, 33, 120-121, 127-131, 134, 143-144, 159, 163, 164

GALE, G., 112n., 201
GALILEO, 25
GALLUP, G., 99n., 202
GALLWAY, T., 58, 202
Jugador, falacia del, 26n.
Juego, 83, 84
de la vida, 23, 88
contra la naturaleza, 84
semántica de la teoría de juegos, 85n.
jugar juegos, 194
GAVAGAI, 31
GAZZANIGA, M., 56n., 202
Contador Geiger, 94n., 135, 139-140, 156
Genes, 38, 61, 72n., 112, 141, 170
Generadores de intuición, la maestría de los, 22, 30-31
Genético,
determinismo, 47n., 61
variación, 88
Ginebra, 110
Célula germinal, división de la, 141
Máquina, el fantasma en la, 19, 19n.
GIBBS, J. P., 181n., 202
GIBSON, J. J., 53, 202
Glándulas como causas, 46
Metas, 38, 79, 130, 144
en conflicto, 104
no hay metas en la evolución, 78
Dios, 34, 35, 38, 93, 103, 137, 157, 176, 188, 189, 194

punto de vista de, 35, 120-121, 143, 159
biólogo semejante a, 24
herramienta de, 22
GÖDEL, K., 44n., 50n., 172
Golden Gate, 123-124
GOLDWATER, B., 145, 147
Golfista, 29, 126, 131, 191
putt de Austin, 167-168
GOMBERG, P., 186, 202
Bueno, 36, 37, 178
GOODMAN, N., 31, 58, 202
Gourmetoscopio, 86
Error gramatical, 25
Gratitud, 28n., 176
Gravedad, 70, 133, 147
GREENSPAN, P., 67n., 78n., 81, 202
GROSS, H., 181n., 183n., 202
Grupo, selección de, 38n.
Grue-bleen, enigma de, 31
Intereses, guardián de, 37, 39
Guerrillas, guerra de, 53
Culpa, 33, 175-180, 186-190
GUNDERSON, K., 96

Pensamiento, hábitos de, 18, 146, 163
HACKING, I., 169n., 202
Parada, problema de la, 46
HAMPSHIRE, S., 105, 114n., 136n., 202
Discapacitado mental, 179
Ahorcado, paradoja del, 13
Suceder. *Véase* hacer *vs* suceder.
Hara kiri intelectual, 126
Determinismo duro, 101
HART, H. L. A., 181n., 202
Harvard, 141
Odio, 27
HAUGELAND, J., 43n., 202
Rumor, 82
Cielo, 181
Heurística, 87, 107, 138, 160-161
HINTIKKA, K. J. J., 85n., 202-203
Historia,
curso de la, 143-146
inevitabilidad histórica, 143, 148n.
HITLER, A., 176, 189
HOBART, R. E., 149, 163n., 203
HOBBES, T., 28, 31, 34, 35, 49-50, 60, 61, 203
HOFSTADTER, D. R., 23-25, 31n., 43-46, 50n., 57n., 72n., 81, 83n., 92n., 99n., 118, 129-130n., 144n., 173n., 203
HOLMSTROM, N., 67n., 203
Homeostasis, 36
Homo sapiens, 100, 182
HONDERICH, T., 203
HONORÉ, A. M., 168, 184n., 203
HOOK, S., 112n., 203
HOSPERS, J., 112n., 204
HOYLE, F., 72n.
Humanista, 110n., 194
HUME, D., 39n., 48, 91, 95n., 146, 166, 204
HUMPHREY, N., 54n., 58n., 204
HUTCHINGS, E., 71n., 204
Hylefobia, 109
Hipnotismo, 18, 21, 30
el terrible hipnotizador, 21

"Yo", 105n., 108, 109
Idea,
de la sensibilidad y de la reflexión, 51
actuar bajo la idea de libertad, 120-150, 191
Ignorancia, 179
del futuro, 121
de la ley no es excusa, 183-184, 193
Ignorar, 87n., 127-128, 132, 162, 191
Illinois, Universidad de, 137n.
Ilusión, 17, 51, 94, 134, 139, 150, 153, 168, 193
en el certamen de arrojar la moneda, 112
de control, 67, 68, 93
del libre albedrío, 28, 191-192
el yo como, 27
irresistible, 29
del motor semántico, 50
estable, 47, 53, 63
valor de la, 178
Imaginación, 19, 25, 29, 36, 47-50, 55, 65, 109, 133, 144, 146, 148-149, 159, 192-193
manejo de la, 133
Imperfección, 33, 46, 53, 59, 156, 178, 181, 187

guardianes imperfectos de intereses, 37
de los motores semánticos, 128n.
Descarte implícito, 118
Esfera contrafáctica implícita, 130n.
Imposibilidad, 126n., 153, 166-170, 187
de los motores semánticos, 43
e inconcebilidad, 63-64, 193
lógica, 161n.
Incompatibilismo, 67, 121, 163n., 165-166
Teorema de incompletud de Gödel, 50n.
Inconcebilidad, 102n.
e imposibilidad, 63-64, 193
Indeterminismo, 65, 94n., 102-103, 144, 155-156, 159, 163n., 193
como secreto terrible, 27
de la mecánica cuántica, 83. *Véase también* Mecánica cuántica.
Referencia indexical, 99n.
Indiferencia, 34n.
del entorno, 75, 78, 138
Individualidad, 87
Razonamiento inductivo, 26n.
Inevitable, 122, 142-150
la guerra nuclear como, 28
Inexorabilidad, 145, 148n., 149
Niños pequeños, 179; *véase también* Niños.
Infinitud, 65, 173, 179, 194
humana, 48n.
Información,
contenido de, 173, 185
recolección de, 48, 127-128, 130, 137-138, 161
la incompresibilidad de la información como azar, 84, 138, 171-172, 173
insensibilidad a la, 109, 124
sobre la ley, 183
procesamiento de, 43, 52, 56, 58, 94, 160
teoría de la, 76n.
Inclinaciones innatas, conocimiento de las, 37, 61, 99n.
Insania, 25-26, 179, 184n.
defensa por, 179-184
Insecto, 23, 26, 54, 83n. *Véase también Sphex.*
Pájaros insectívoros, 128, 170
Submetas instrumentales, 36
Inteligencia, 53
Intencionalidad, 53
actitud intencional, 78, 129n.
Intenciones, 79-81
Interaccionismo, 19n.
Intereses, 35-40, 59, 74, 78-80, 125, 128, 135, 138
Interpersonal,
actitud, 49
relaciones, 63
Interpretación de las personas, 163
Introspección, 26, 92, 96, 157n.
prueba del tornasol, 50
Intuición, 18, 28, 52, 57, 93, 152, 154, 173, 180
generador, 25, 30-31, 48-49, 52, 57, 59n., 64, 81, 97, 124, 139, 150, 152, 192-193
Invisible,
senderos causales, 95
mano, 78
carcelero, 20, 25, 124
Irracionalidad, 46, 185; *véase también* Insania;
miedo de volar, 153
Iteración, 44-45

JACOB, F., 120, 127, 204
Carcelero, 182; *véase también* Prisión;
invisible, 20, 25, 124
caprichoso, 136
JAMES, W., 152, 204
JAUCH, J. M., 173n., 204
JAYNES, J., 58n., 86n., 100n., 106n., 204
JENNINGS, P., 78n., 204
JENNINGS, R. E., 171n., 204
Jueces, 181-182, 183n.
Jurado, 120
"Sólo un cuento", 54, 58, 60, 64, 176
Justicia, 183-184
Justificación por medio de razones, 47

KAHNEMANN, D., 26, 204
KANT, I., 32-35, 39, 60, 65, 87, 108, 152, 179, 187, 189, 204
Mantener la cabeza baja, 29, 126, 132
KIERKEGAARD, S., 120, 127
Conocimiento, 183
"conocer" y "conocimiento", 54-55, 183
teoría causal del, 74n.
de la causalidad, 81
KREBS, J., 41, 204
KRISTOFFERSON, K., 142
KYBURG, H., 26n., 204

Lenguaje, 44, 65, 119
evolución del, 58n.
aprendizaje del, 114
público, 47, 57, 59-61
LAPLACE, P., 64, 66, 77, 83, 120, 129, 173, 204
Ley, 179, 186
secreta, 184
que nos damos a nosotros mismos, 108
LEDOUX, J., 56n., 204
LEIBNIZ, W., 87
LEM, S., 22, 204
"Suéltese", 86-90
LEVI, I., 133n., 205
LEWONTIN, R., 129n.
Ley, credibilidad de la, 181-185
Libertarianismo, 74, 93n., 95n.
Luz, velocidad de la, 71, 169
Lingüística, 43
Vínculo, 162
LOCKE, J., 35, 51-52, 104, 205
Locomoción, 36
Sofistería lógica, 14
Posibilidad lógica, 141, 169
Londres, 27, 91, 120-121
el gran terremoto de Londres de 1981, 145, 147
Mirar hacia adelante, 29-30, 120
Examinar con excesivo detenimiento, 26, 185, 187
Circuito, 44, 67; *véase también* Retroalimentación.
"Una cierta soltura", 166
Lotería, 139-140
Amor, 17, 43, 124-125, 191
LUCAS, J. R., 44n., 48n., 50n., 67n., 148, 205
Suerte, 98, 109-118, 140-141, 168, 185
LUCRECIO CARO, T., 14, 140, 205
LUTERO, M., 153, 155, 163n.
LYCAN, W. G., 99n.

MACINTYRE, A., 46, 205
MACKAY, D. M., 77n., 205
MACKIE, J. L., 159, 205
MACNAMARA, J., 47
Macroscópico contra microscópico, 94n., 133-134, 156, 162, 173
Loco,74, 153, 179; *véase también* Insania;
locura como libertad, 35, 74
locamente enamorado, 125
Magdalen College, 19n.
Poder mágico, 100-101
Línea Maginot, 53
Maine, 20
Establecer una diferencia. *Véase* Diferencia.
Hacer verdadera una predicción, 23
Entorno maléfico,75; *véase también* Indiferencia.
Mamíferos, 102
Maniqueo, 78
Imagen manifiesta, 130, 133, 142, 150
Manipulación,
inmunidad a la, 89
solapada, 53, 81
MANSON, Ch., 106
Mecánica cuántica, interpretación de los múltiples mundos de la, 144n., 159
Maratón, 114
Error, margen de, 79, 89
Deliberador determinista Mark I, 135, 139, 142, 160
Deliberador aleatorio Mark II, 139
Marte, 71, 73-74, 145
Marcianos, 85, 160
Masa, término de, 128

"agua" como, 134
Materialismo, 13-14, 24, 35
Matemática, 172
Apareamiento, 36
Preocupación, 17, 18, 28, 142, 157, 160, 178
Máximas, búsqueda de, 80n.
MCCAWLEY, J. D., 13
MCDERMOTT, D.,132, 205
Significado, 45
la recurrencia eterna y el, 22
captación del, 41
de nuestra vida, 177
no es una propiedad física, 41-42
transductores de, 41-54
Mecanismo, 23, 35, 41-45, 99, 195
Memoria, 187
MENÓN, 31, 55n.
Mental,
deficiente, 61
"mental" como sinónimo de estú pido o demente, 47
Sirenas, 166
MERSENNE, M., 34n.
Meta-,
ética, 62
"actividades meta-", 44
lenguaje, 44, 106, 108
actividad de metanivel, 82
control de metanivel, 82, 103-105, 109
Metáfora, 14, 19, 78-79, 100, 192
Metafísica, 18, 30, 98n., 108, 142, 148, 156, 159, 175, 180, 192, 195
M15, 38
Descripción microfísica y de micro–nivel, 109, 129, 131, 135, 138, 157, 160-162, 168, 184
MILGRAM, S., 26n., 205
MILL, J. S., 70, 98n., 108, 205
MILLER, D. S., 149n.
Imitación, 70
Mente, 42, 50, 56
filosofía de la, 52
leer la, 23, 82
Milagro, 145
Lógica modal, 143, 170, 171n.
MONOD, J., 35n., 75n., 171
Monstruo, 180, 189
Moral,
acto, 176-179
mundo de los conceptos morales, 176-179, 190
decisiones, 116
educación, 114
fuerza, 163
visiones, 107
Moralidad, 185
origen de la, 60, 61, O., 80n., 205
MORGENSTERN, O., 80n., 205
Madre Naturaleza, 39, 78, 89, 127, 147, 173
Motivación,
de los ciudadanos, 182
de proyectos filosóficos, 16, 19, 23, 30
Ser movido por razones, 40
MOZART, W., 26, 92, 97
Asesinato premeditado, 122
Asesino, 118, 152, 154-155, 188
Mutación, 139, 171

NAGEL, T., 31n., 47, 50n., 92, 111, 116, 205
NAPOLEÓN, 20
Narratización, 106n.
NASA, 71, 74
Naturalismo, 65, 93, 98, 112, 180, 188, 193-194
Naturaleza, 64-65, 88, 149
no juega a las escondidas, 137
leyes de la, 169
estado de, 31, 60-61, 184
modo de la, 60n.
Necesidad, 166, 170
causal, 143, 148, 150
libertad como conciencia de la necesidad, 82n.
lógica modal de la, 143
selectiva, 170
"Saber lo estrictamente necesario", principio de, 38-39, 54
Cirujano nefando, el, 20, 21, 81, 152
NEISSER, U., 80n., 206
Neurona, 92
NEWTON, I., 129n., 146, 173

Nueva York, puerto de, 99n.
Cataratas del Niágara, 147
NIETZSCHE, F., 22, 59, 62, 91, 98, 106, 176, 181, 206
Nihilismo, 175-179
NIXON, R., 176, 180, 189
Noblesse oblige, 113
Ruido, 75n.
Imposibilidad nomológica, 161n., 169
Normal, inercia de lo, 146-147
advertir, 48, 55, 82, 158
Ahora disparado hacia el futuro, 121
NOZICK, R., 21, 29n., 65-68, 72n., 74, 81, 87n., 109, 149, 158n., 206
Guerra nuclear, 28

Objetividad científica, 27
Objet trouvé, 103
Olvido, 38
Obscurum per obscurius, 109
Edipo, 123
Compromiso ontológico, 147
Abierto,
al agente, 79n., 132
comunidad abierta, 136
futuro, 121, 132, 134, 159, 193
Condicionamiento operativo, 26
Investigación operativa, 162n.
Oportunidad, 27, 45, 82, 88-90, 113, 115, 118, 121-122, 134-142, 146, 150, 173, 191, 194
rasa, 136
de robot, 135
Soluciones óptimas, 40-41, 54, 87, 128, 133, 161-162, 165, 185-186
Opciones,
disponibles, 89-90, 121
mantener las opciones abiertas, 142
Lenguaje ordinario, filosofía del, 68
Organismos,
los seres humanos como, 24
seres humanos en competencia con, 23
ORGEL, segunda regla de, 38n.
Posición original, 31
Pecado original, 187
OSWALD, L. H., 147
Otras mentes, 95
Sobredeterminación, 153
Simplificación, exceso de, 98; *véase también* Simplicidad.
Oxford, 19n., 27, 29n., 40, 47n., 71
Oxígeno, 166, 173

Dolor, 17, 36, 96
PANGLOSS, 165
Caso paradigmático, 68n.
Parálisis, 18, 27, 30, 97
Parodia, 22
Actitud participativa, 28n., 63
Parus major, 40
Pasión, 39n.
Pasado, 66-67, 89, 101, 121, 128n., 140, 143, 149, 159, 163, 195
Pautas,
en la actividad, 84, 88, 103, 138, 148, 170
inescrutabilidad de las, 83, 138, 172
reconocimiento de, 36, 44, 50, 52
Zonzo, 118
PEARS, D., 131, 206
Perfección, 44n., 46, 65, 149, 171, 187, 189
perfeccionismo ingenuo, 55
Periodicidad,85n., 172n.
Movimiento perpetuo, máquina de, 43
"Podría haber obrado de otra manera", 29, 33, 48n., 79n., 151-174, 193
Personalidad, 87
Persona moral, 98-99, 106, 114
Perspectiva,
de la primera persona, 95
estrecha del agente, 121
Persuasión, 62
Fenomenología, 95n.
Filósofos, 14-18, 40, 42, 44n., 46-47, 51, 55n., 59, 67, 80, 109, 132, 140, 148, 157, 167, 169, 179, 192
Método filosófico, 15-18, 30, 31-32, 108
Filogenia, 44
Condiciones y procesos físicos, 42-44, 52

Físicos, 193
Física, 88, 121, 157, 166, 169, 173.
Véase también Microfísica y mecánica cuántica.
PICASSO, P., 124
Avión, piloto de, 79, 82, 103
Planificación, 133, 135, 147, 160, 168-169, 190
PLATÓN, 31, 55n.
Formas platónicas, 38
Juguetes, 21, 22, 30
POINCARÉ, H., 26n.
Punto de vista, 37, 170; *véase también* Perspectiva.
Libertad política, 17
Polyana, 78
Pontryagin, principio del máximo de, 41
POPPER, K., 77n., 131, 206
PORTER, C., 161
Posibilidad, 84, 120-121, 129-133, 136n., 137, 141, 152, 160, 165-171, 173, 193
epistémica, 129, 169, 173
lógica, 141, 165, 167
seria, 133
Mundos posibles, 162n.
el mejor de los, 165
Potencialidad, 129, 168, 169n.
Pobreza, 192
Poder, 30, 33, 170
POWERS, L., 55n., 206
Predador, 39, 170
evitar el, 36, 170
saturación, 39
Predecibilidad, 77, 83, 169n., 170
Predicción, 23, 127-130, 137, 163, 173
Prevención, 143-148, 194-195
Prisión, 18, 19-21, 25, 30, 181-182, 195
prisionero, 122, 136
dilema del prisionero, 106, 107n.
Privacidad, 57, 79, 188
Acceso privilegiado, 157n.
Probabilidad, teoría de la, 80n., 85n., 166-167, 174
Problemas, resolución de, 40-41, 124, 130-131, 138
Programación, 40-41, 154
Propiedad proyectable, 111
Proyectos, 17, 104, 119, 192
importancia de nuestros, 17
de los árboles, 123
PRUFROCK, J. Alfred, 33
Seudoaleatorio, 84, 85n., 94n., 138-140, 161, 162n., 172, 173
Psicoanálisis, 52, 58
Psicología, 47, 95n.
psicólogos, 26, 80n.
causalidad "psicológica", 45-46, 61-62
Ley, publicidad de la, 183
Publishers' Clearing House, 140
Títeres, juego de, 67
Castigo, 13, 176-184, 185, 188, 190
Marioneta, 18, 21, 30, 67, 195
no robot, 71
Propósito, 134
del cerebro, 42-43
consciente, 52
PUTNAM, H., 31, 206
PYNCHON, T., 93, 206
Evaluación cualitativa y no cualitativa, 104n.
Mecánica cuántica, 83-84, 94n., 109, 139-141, 144n., 156, 159, 165, 172
Colas, teoría de las, 162n.

QUINE, W. V. O., 31, 128n., 162, 206

RAAB, F. V., 157n., 206
RABIN, M., 172, 206
Radar, 99n.
Radical,
acción, 86
elección, 107
libertad, 82, 163n.
Azar, 83, 86n., 94n., 111, 112, 119, 123, 128, 140, 149, 162n., 172
Véanse también Seudoaleatorio; Mecánica cuántica.
Algoritmos, 172
Sentido matemático, 84
Número aleatorio, generación del, 84, 139-140, 161, 172
Atomos, desvío de, 14, 139
Racional,

animales, 32, 182
reconstrucción, 181
voluntad, 42, 43, 65, 179, 189
Diseño, fundamento lógico de un, 38, 126-127, 132, 176, 180-181, 185
flotante, 38, 82n., 83, 88, 176
Racionalismo, 87
Racionalidad, 34-65, 119, 123, 126, 127, 134, 181-184, 191-194
a-racionalidad, 46
perfecta, 34-35
RAWLS, J., 206-207
Reacciones a reacciones, 44, 59
Actitudes reactivas, 28n.
Reader's Digest, 140
REAGAN, R., 71, 164
Razón, 34-65, 123, 125
aparente, 40
razones como *abstracta,* 38, 59
dictados de la, 34-35, 38-41, 45, 87, 106-107n., 149, 153
las razones se evaporan, 27
incapaz de, 117
pura práctica, 40
los acontecimientos aleatorios ocurren "por ninguna razón", 14
razonamiento, 160, 187
razones y causas, 42, 46
voz de la, 109, 125
Reconocimiento,
de intereses, 59
de significados, 41
de oportunidades, 135-136
de razones, 49, 50
Reconciliacionismo, 14
Reductio ad absurdum, 103, 166
Reflexión, 105-106
Reflexividad, 44-46, 50-60, 62, 191
Remordimiento, 160, 163n., 178, 189
REID, T., 93, 207
MOR (movimientos oculares rápidos), 137
Control remoto, 69-72, 95n., 108
Replicante, 35-36, 141
Representación mental, 44-45, 61
Presentar razones a alguien, 49-50, 59, 61, 163, 190
Autorreproche, 188
Res cogitans, 27
Resentimiento, 27, 28n.
Resignación, 14, 27-28, 122, 124
Resistencia, 107
Resistencialismo, 78n.
Responsabilidad, 46-47, 65, 79n., 92-94, 96, 101, 108-113, 117-118, 151-157, 163-165, 175-177, 179-180, 185-191, 194
disminuida, 33, 179-184
considerar responsables a los de más, 33, 176-179
asumir, 164, 192
Ressentiment, 176
Retardo mental, 25, 47, 61
RIDLEY, tortugas marinas de, 39
Bien y mal, origen del, 60
Hacer lo correcto, 40, 87, 161
Rigidez, supuesto de, 129n.
RILKE, R. M., 86, 89
Riesgo, 185, 188
ROBBINS, T., 22, 49, 207
ROBINSON CRUSOE, 80n.
Robot, 55, 71-72, 74, 135, 139, 142-143, 160-162
oportunidad de, 135
Solidez, 163, 192
Piedra, papel y tijeras (juego), 22-23, 83
Desempeñar un papel, 14
RORTY, A., 100n., 125, 207
RORTY, R., 50n., 63, 207
Ruleta, rueda de la, 77
RUBY, J., 147
Rusia, 110
RYLE, G., 17-19

SAARINEN, E., 85n., 207
Saccades, 137n., 137n.
SANFORD, D., 103n., 207
Cordura, 179-180; *véase también* Insania.
Santa Claus, 23
SARTRE, J. P., 100, 107, 133, 189, 207
SCHANK, R., 147
SCHIFFER, E., 125, 207
SCHNEEWIND, J., 98n.; 208
SCHOTCH, P. K., 171n., 208

Ciencia, 141, 162, 166, 168
inaccesible para las arañas, 64
perturbadora del universo, 33
no hace verdadero al determinismo, 28
puede decirnos demasiado, 27
la filosofía como semejante a la, 14-15, 31-32
lo que la ciencia muestra, 14-15
Ciencia ficción, 48-49, 137n.
Método científico, 15, 76, 134, 142, 148, 150, 163
Escultura, 15-16, 30, 85n.
volcánica, 36
Escila y Caribdis, 79
Investigación, espacio de, 106
SEARLE, J., 31, 50n., 208
Segunda ley de la termodinámica, 119
Secreto,
agente, 53
terrible, 27, 188-191
leyes secretas, 183
Buscar,
el alimento, 36
el bien, 37
máximas, 80n.
SEIDENFELD, T., 85n., 208
Yo, 26, 33-34, 55-56, 89, 119
activo, 96-97
consciente, 58
amado, 98
que desaparece, 26-27, 91-98
brechas en el conocimiento del, 97
como lugar de autocontrol, 98-100
como punto de vista, 96
que se hace a sí mismo, 91-119, 179, 186, 189, 193
Auto-
elección, 101, 105
condena, 189
conciencia, 52, 57, 100
control, 32, 66-90, 92, 98-100, 103, 112, 117, 119, 138, 141, 146, 165, 168, 187, 191-192, 194
creación, 103-107, 113
crítica, 54
engaño, 27, 33, 125, 180, 188, 191
evaluación, 58, 105-106
perfeccionamiento, 105, 113
conocimiento, 77n., 105, 107
monitoreo, 53-54, 59-60, 72
movimiento, 95n.
predicción, 131
presentación, 188
preservación, 36, 60, 131
rediseño, 45, 109n., 163-164
reparación, 36
réplica, 35-36
reproche, 188
síntesis, 109n.
impredecibilidad, 131-132
observación, 46, 131, 132
SELLARS, W., 130, 208
Motores semánticos, 43-44, 50
Semántica, 43-44
SHANDY, Tristram, 132
SHER, G., 103n., 114n., 208
Barajar (cartas), 77, 85, 88
SIMON, H., 87n., 127n., 129n., 208
Causa, simple contra compleja, 25
Simplicidad,
de los ejemplos, 46-47, 49, 193
de los experimentos, 76
de las tareas de procesamiento de la información, 127-128, 129
Pecado, 175, 179
original, 187
Singularidad, 92
Habilidad, 113-116, 168
SKINNER, B. F., 26, 73-76, 83n., 111n., 176, 177n., 208
Esclavo, 21
Prestidigitación, 22
Deslizamiento, ejes de, 129n.
SLOTE, M., 67n., 143, 165, 166, 170, 180n., 208
SLOVIC, P., 26
SOBER, E., 166n.
Ciencias sociales, 166
SÓCRATES, 179, 189
Sociobiología, 6
Determinismo blando, 101
Sorites, 102
Sortilegio, 86n.
Alma, 92, 190
Espacio-tiempo, gusano del, 159

Especies, 36, 38n., 128, 130, 166-167
Espectador, actitud del, 199
Habla, 56
Sphex ichneumoneus, 23-27, 37-40, 44, 46, 50-51, 54, 59, 63-65, 80, 87, 105, 191
Arañas, 64
SPINOZA, B., 34, 208
Cerebro seccionado, sujetos con el, 56n.
Espontaneidad, 34n., 87, 89
Espía, 34
Soslayar, 162
  en el caso de los escultores, 16
Sistemas estables, 173-174
Stanford, Universidad de, 141
Estado, 68, 129-130, 161-162, 167-168
Estadística, 84, 160n., 166
STIDHAM, S., 162n.
Estímulos, 73-74
  control de los, 74n.
Estoicismo, 14, 20n.
STRAIGHT, S., 58n., 208
STRAWSON, P., 28n., 47n., 52, 63-64, 93n., 100n., 189, 208-209
Luchar, 36-37, 59, 165
Estupidez, 37n., 44, 47, 126, 188
Estilo, 43n., 58, 103-104, 187
Subliminal, 73, 163n.
  imagen de la razón, 39
Sub specie aeternitatis, 144, 159
Metro, 91
Suicidio, 177
Inteligencia superior, 24
Superstición, 111, 123, 126, 140
Suspender,
  la ejecución de un deseo, 104
  el juicio, 96-97
SVENGALI, Dr., 21
Swarthmore College, 141
Guardia Suiza, 182
Interruptores, 94
Motores sintácticos, 43
Sintacticidad, 46
Sintaxis, 43

TAYLOR, C., 104n., 106-108, 209
TAYLOR, R., 126n., 209
Teleología, 35
Teleonomía, 35n.
Terapia, 101
Termostatos, 66
Pensamiento sin pensador, 38
Tercera persona, perspectiva de la, 71
THORP, J., 95n., 209
Experimento mental. *Véase también* Generador de intuición;
  controlado, 81
  dos tipos, 25
Three Mile Island, 147
Tiempo, 13, 120, 144, 159, 190, 193
  presión del, 43n., 86, 104-105, 107-108, 127-131, 134, 138, 183, 187
  real, 131, 134, 137n., 139
  presente sin tiempo, 144
Paros grandes, 40
TOLSTOI, L., 27
Trompo, de madera, 28
Tortura, 24, 154
Tout comprendre c'est tout pardonner, 47
Juguetes, 21-22
Rastreo, 76, 128-129, 132, 134, 149
Tragedia, 14, 123
Transductores, 41-45, 94
Arboles, 123
TRUMAN, H., 93
Verdad, 74, 81, 87, 105
  nos hará libres, 80n.
Probar, 97, 130n., 165, 167-168, 190
Tumor cerebral, 21
Turing, máquina de, 50n.
"Girar la perilla", 81, 139
Tortugas, 39
TVERSKY, A., 26, 209
TWAIN, Mark, 142
Tierra Gemela, 31
Bandido, de los dos brazos, 40
Tyché, 165
Aldea tirolesa, 22, 49
ULLMAN, S., 129n., 209
Incertidumbre, 79, 89, 129, 132n., 169n.
  amplificador de, 106-107n.
Inconsciente, 51-53

Argumento subterráneo, 42
Acceso subprivilegiado, 96
Vida sin examen, 105
Injusto, 113-114, 139-140, 178, 185, 189
Unicornios, 166-167, 169
Estados Unidos, 110
Universo,
  perturbar el, 33, 144, 147-148
  nuestro lugar en el, 120
  replicantes en el, 36
Desenmascarar, 24
Motor inmóvil, 93, 97, 189
Impredecibilidad, 23, 169n., 173-174
Influencia inconsciente, 21, 81
"Depender de" alguien, 33, 132, 138, 153, 193

VALÉRY, P., 120
Valor, 62, 119
  creación de, 108
  juicios de, 178
  revaloración de, 105-106
  valoración de, 62
VAN INWAGEN, P., 67n., 102n., 123n., 124, 151, 169n., 209
VENTRE, P. M., 78n.
VESCO, R., 189
Viking Autónomo, 71n., 89
Violación, 155
Virtud, 34
Voz de la razón. *Véase* Razón.
Volición, 39n., 95n.
Voluntario, 94-97
VON NEUMANN, J., 80n., 209
VON WRIGHT, 146n., 149, 209
Apuesta, 23, 185
Avispa, 23-27, 30, 66; *véase también Sphex ichneumoneus.*
WATSON, G., 93, 209
Paquete de ondas, colapso del, 109
WAYNFLETE,
  conferencias, 19n.
  profesor, 19n.
Occidental, el libre albedrío como un problema, 17
Fortuna, rueda de la, 14
WHITAKER, H., 26, 209
WIENER, N., 48n., 83n., 128n., 160n., 162, 165, 209
WIGGINS, D., 67n., 74, 143, 209
Voluntad, 35
  poder de la, 97
  racional, 41, 43, 65, 179, 189
  falta de, 125, 188
WILLIAMS, B. A. O., 39n., 111n., 113, 209-210
WIMSATT, W., 83n., 113n., 128n., 129n., 162, 172n., 210
Sentido común, 103, 114, 142
Deseo, expresión de, 102, 126, 135, 191-193
WITTGENSTEIN, L., 28, 32, 42, 210
WOLF, S., 39n., 177n., 179, 187, 210
WOLFF, R. P., 210
WOOLDRIDGE, D., 24, 210

ZAITCHIK, 103n., 210
Zombie, 21, 154

NOTA FINAL

Le recordamos que este libro ha sido prestado gratuitamente para uso exclusivamente educacional bajo condición de ser destruido una vez leído. Si es así, destrúyalo en forma inmediata.

Súmese como voluntario o donante y promueva este proyecto en su comunidad para que otras personas que no tienen acceso a bibliotecas se vean beneficiadas al igual que usted.

"Es detestable esa avaricia que tienen los que, sabiendo algo, no procuran la transmisión de esos conocimientos".

—*Miguel de Unamuno*

Para otras publicaciones visite:
www.lecturasinegoismo.com
Facebook: Lectura sin Egoísmo
Twitter: @LectSinEgo
o en su defecto escríbanos a:
lecturasinegoismo@gmail.com
Referencia: 779

Borre el libro una vez leído

Ciencias Cognitivas

LA *LIBERTAD* de *ACCION*

El examen del problema del libre albedrío muestra que muchas de las profundas dudas acerca de su posibilidad no son más que preocupaciones académicas con poca o ninguna relación con la conciencia de nuestros actos, elecciones y posibilidades.

Las normas de la interacción social, las instancias ajenas y subjetivas de control de la conducta y, sobre todo, el hecho fundamental de que toda elección implica la exclusión de otra son fenómenos que dejan un margen bastante estrecho a nuestra libertad. Pero, ¿tiene sentido que nos desesperemos por no poder anular el pasado, por no poder prevenir ciertos acontecimientos presentes o por no poder crearnos de la nada ni elegir dos alternativas a la vez?

En este texto provocador y refrescante, **Daniel C. Dennett** desmitifica y reafirma las posibilidades de la libertad, subrayando tres condiciones básicas de toda decisión: ser libre, racional y responsable. La necesaria conexión de estos tres términos deja claro que lo que está en juego no es tanto la libertad sino el hecho de si la deseamos realmente.

De este mismo autor, Editorial Gedisa ha publicado también *La actitud intencional*.

ISBN 84-7432-426-2

9 788474 324266

Código : 2.368

gedisa editorial

Colección Hombre y Sociedad
Serie
CLA·DE·MA